

# BURGESS BOUNDS FOR MULTI-DIMENSIONAL SHORT MIXED CHARACTER SUMS

L. B. PIERCE

**ABSTRACT.** This paper proves Burgess bounds for short mixed character sums in multi-dimensional settings. The mixed character sums we consider involve both an exponential evaluated at a real-valued multivariate polynomial  $f$ , and a product of multiplicative Dirichlet characters. We combine a multi-dimensional Burgess method with recent results on multi-dimensional Vinogradov Mean Value Theorems for translation-dilation invariant systems in order to prove character sum bounds in  $k \geq 1$  dimensions that recapture the Burgess bound in dimension 1. Moreover, we show that by embedding any given polynomial  $f$  into an advantageously chosen translation-dilation invariant system constructed in terms of  $f$ , we may in many cases significantly improve the bound for the associated character sum, due to a novel phenomenon that occurs only in dimensions  $k \geq 2$ .

## 1. INTRODUCTION

Let  $\chi(n)$  be a non-principal multiplicative Dirichlet character to a modulus  $q$ , and consider the character sum

$$(1.1) \quad S(N, H) = \sum_{N < n \leq N+H} \chi(n).$$

The Pólya-Vinogradov inequality states that

$$S(N, H) \ll q^{1/2} \log q,$$

which is nontrivial only if the length  $H$  of the character sum is longer than  $q^{1/2} \log q$ . Burgess famously improved on this in a series of papers [3] [4] [5] [7], proving (among more general results) that for  $\chi$  a non-principal multiplicative character to a prime modulus  $q$ ,

$$(1.2) \quad S(N, H) \ll H^{1-\frac{1}{r}} q^{\frac{r+1}{4r^2}} \log q,$$

for any integer  $r \geq 1$ , uniformly in  $N$ . This provides a nontrivial estimate for  $S(N, H)$  as soon as  $H > q^{1/4+\varepsilon}$ ; more precisely if  $H = q^{1/4+\kappa}$ , then the Burgess bound is of size  $Hq^{-\delta}$  with

$$(1.3) \quad \delta \approx \kappa^2.$$

The Burgess bound found immediate applications in an upper bound for the least quadratic non-residue modulo a prime and a celebrated sub-convexity estimate for Dirichlet  $L$ -functions, and has since been used in a wide range of problems in analytic number theory. Burgess's original strategy has also been refined and simplified (for very recent examples see [10] [12]) and adapted to other problems (for example [11] [15]), but its main utility

---

2010 *Mathematics Subject Classification.* 11L40 (11P05, 11D45).

*Key words and phrases.* Burgess bound, short character sum, Vinogradov Mean Value Theorem.

currently remains limited to a few types of short character sums. It would be highly desirable to generalize the Burgess method further to a wide range of character sums involving additive and multiplicative characters, polynomial arguments, and multiple dimensions.

In the present work we develop Burgess bounds for multi-dimensional short mixed character sums of the following form. For each  $i = 1, \dots, k$ , let  $\chi_i$  be a non-principal multiplicative character modulo a prime  $q_i$ . Let  $f$  be a real-valued polynomial of total degree  $d$  in  $k$  variables and set

$$S_k(f; \mathbf{N}, \mathbf{H}) = \sum_{\substack{\mathbf{x} \in \mathbb{Z}^k \\ \mathbf{x} \in (\mathbf{N}, \mathbf{N} + \mathbf{H}]}} e(f(\mathbf{x})) \chi_1(x_1) \cdots \chi_k(x_k)$$

for any  $k$ -tuple  $\mathbf{N} = (N_1, \dots, N_k)$  of real numbers and  $k$ -tuple  $\mathbf{H} = (H_1, \dots, H_k)$  of positive real numbers, where

$$(\mathbf{N}, \mathbf{N} + \mathbf{H}) = (N_1, N_1 + H_1] \times (N_2, N_2 + H_2] \times \cdots \times (N_k, N_k + H_k]$$

denotes the corresponding box in  $\mathbb{R}^k$ , with volume  $\|\mathbf{H}\| := H_1 \cdots H_k$ . Note that we do not assume the primes  $q_i$  are distinct, and in particular an interesting special case arises when all the  $q_i$  are equal to a fixed prime  $q$ . To avoid vacuous cases we always assume  $f$  has positive degree with respect to each of the  $k$  variables, and that  $H_i \geq 1$  for  $i = 1, \dots, k$ .

We note the trivial bound

$$(1.4) \quad S_k(f; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|.$$

Nontrivial upper bounds for  $S_k(f; \mathbf{N}, \mathbf{H})$ , particularly when  $H_i$  is “short” relative to  $q_i$ , are expected to have a variety of applications, for example to counting integral points on certain hypersurfaces, such as multi-dimensional generalizations of the Markoff-Hurwitz and Dwork hypersurfaces (see related work [16], [9]).

We will prove bounds that are nontrivial when  $H_i \gg q_i^{1/4+\varepsilon}$  by developing a multi-dimensional version of the Burgess method that allows us to apply recent results of Parsell, Prendiville and Wooley [14] on multi-dimensional Vinogradov Mean Value Theorems. The basic framework of this approach is inspired by [13], which treats the one-dimensional case, but a new phenomenon arises in dimensions  $k \geq 2$ . To make this phenomenon clear, we focus now on two specific results which we may frame in very concrete terms. (Both are immediate corollaries of our most general result, Theorem 2.1, which is stated in terms of translation-dilation invariant systems; see Section 2.)

The key strategy of our multi-dimensional Burgess method will transform the original sum  $S_k(f; \mathbf{N}, \mathbf{H})$  into a collection of many shorter sums  $S_k(\tilde{f}; \tilde{\mathbf{N}}, \tilde{\mathbf{H}})$  with other polynomials  $\tilde{f}$  and tuples  $\tilde{\mathbf{N}}, \tilde{\mathbf{H}}$ . The transformations  $\tilde{f}$  of  $f$  will live inside a certain family, which we may choose to construct in various ways. If we embed  $f$  into the family of all polynomials in  $k$  variables of degree at most  $d$ , we obtain a direct generalization of the work of [13] to  $k$  dimensions (Theorem 1.1). But a more sophisticated embedding of  $f$  into a potentially much smaller family of polynomials allows us to obtain a sharper result (Theorem 1.2). We now describe these two results.

**1.1. Generic embedding.** We suppose we are given a fixed real-valued polynomial  $f$  of total degree  $d$  in  $k$  variables, and a corresponding sum  $S_k(f; \mathbf{N}, \mathbf{H})$ . For  $\mathbf{x} \in \mathbb{Z}^k$  we will use multi-index notation, so that for a tuple  $\beta = (\beta_1, \dots, \beta_k) \in \mathbb{Z}_{\geq 0}^k$  we have  $\mathbf{x}^\beta = x_1^{\beta_1} \cdots x_k^{\beta_k}$ . We let  $|\beta| = \beta_1 + \cdots + \beta_k$  denote the total degree of the monomial  $\mathbf{x}^\beta$ . We consider the

system of Diophantine equations given by

$$(1.5) \quad \mathbf{x}_1^\beta + \cdots + \mathbf{x}_r^\beta = \mathbf{x}_{r+1}^\beta + \cdots + \mathbf{x}_{2r}^\beta, \quad \text{for all } 1 \leq |\beta| \leq d$$

where each  $\mathbf{x}_j \in \mathbb{Z}^k$ . We let  $R_{d,k}$  denote the number of equations in this system and  $M_{d,k}$  denote the sum of the total degrees appearing in the system; we recall that

$$(1.6) \quad R_{d,k} = \binom{k+d}{k} - 1, \quad M_{d,k} = d \binom{k+d}{k} \frac{k}{k+1}.$$

Let  $J_{r,d,k}(X)$  denote the number of solutions to the system (1.5) with  $1 \leq x_{j,i} \leq X$  for all  $1 \leq j \leq 2r$ ,  $1 \leq i \leq k$ . The main conjecture in the setting of multi-dimensional Vinogradov Mean Value Theorems is that for all  $r$  sufficiently large with respect to  $d$  and  $k$ ,

$$(1.7) \quad J_{r,d,k}(X) \ll X^{2rk - M_{d,k} + \varepsilon}.$$

In the case  $d = 1$ , (1.7) holds trivially for all  $k, r \geq 1$ . The case of  $d \geq 2$  is highly nontrivial. Nevertheless, recently Parsell, Prendiville, and Wooley [14] have proved this for a nearly optimal range of  $r$ :

**Theorem A** (Theorem 1.1 of [14]). *For  $k \geq 1$ ,  $d \geq 2$ , if  $r \geq R_{d,k}(d+1)$ , the main conjecture (1.7) holds for every  $\varepsilon > 0$ .*

We combine this with the Burgess method to prove our first result; here we use the conventions that  $\mathbf{q} = (q_1, \dots, q_k)$ ,  $\|\mathbf{q}\| = q_1 \cdots q_k$ ,  $q_{\max} = \max\{q_i\}$ ,  $q_{\min} = \min\{q_i\}$ .

**Theorem 1.1.** *Let  $M = M_{d,k}$ . For any  $r > R_{d,k}(d+1)$ , if  $q_i^{\frac{1}{2(r-M)}} < H_i < q_i^{\frac{1}{2} + \frac{1}{4(r-M)}}$  for each  $i = 1, \dots, k$ , then*

$$S_k(f; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1 - \frac{1}{r}} \|\mathbf{q}\|^{\frac{-r-M+1}{4r(r-M)} + \frac{M}{4kr(r-M)} + \varepsilon} q_{\max}^{\frac{2rk}{4r(r-M)}} q_{\min}^{-\frac{M}{4r(r-M)}},$$

*uniformly in  $\mathbf{N}$ , with implied constant dependent on  $r, d, k, \varepsilon$  and independent of the coefficients of  $f$ .*

It is illustrative to record the result when all the moduli  $q_i$  are equal:

**Corollary 1.1.1.** *Under the hypotheses of Theorem 1.1, if in addition  $q_1 = \cdots = q_k = q$  for a fixed prime  $q$ ,*

$$S_k(f; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1 - \frac{1}{r}} q^{\frac{k(r+1-M)}{4r(r-M)} + \varepsilon}.$$

In general, for any dimension  $k$ , we may check the strength of Corollary 1.1.1 as follows: it is nontrivial if each  $H_i = q^{1/4 + \kappa_i}$  for  $\kappa_i > 0$ , in which case choosing  $r$  optimally shows that Corollary 1.1.1 provides a bound of size  $\|\mathbf{H}\|q^{-\delta}$ , where

$$\delta \approx \frac{\left(\sum_{i=1}^k \kappa_i\right)^2}{k}.$$

Thus the improvement in the bound  $\|\mathbf{H}\|q^{-\delta}$  over the trivial bound is independent of the degree  $d$  of the polynomial  $f$ , and recovers the Burgess result (1.3) when  $k = 1$ . Moreover, Theorem 1.1 recovers Theorem 1.3 of [13] in dimension  $k = 1$ .

Thus this is a natural multi-dimensional generalized Burgess bound for  $S_k(f; \mathbf{N}, \mathbf{H})$ . However, in dimensions  $k \geq 2$  another effect can come into play, which we now describe.

**1.2. Minimal embedding.** Given a polynomial  $f$  of total degree  $d$  in  $k$  variables, we can write it in terms of its coefficients  $f_\beta$  as

$$f(\mathbf{x}) = \sum_{\beta \in \Lambda(f)} f_\beta \mathbf{x}^\beta,$$

where  $\Lambda(f)$  is the set of nonzero multi-indices corresponding to monomials in  $f$  with non-zero coefficients. (Since the size of  $|S_k(f; \mathbf{N}, \mathbf{H})|$  is unaffected by any constant term in  $f$ , we may assume that  $f$  has no constant term.) Next, we construct a set comprised of all the distinct non-constant monomials (rescaled to be monic) that appear in partial derivatives of  $f$  of any order; we will call this set  $\mathbf{F}(f)$ . To construct  $\mathbf{F}(f)$  explicitly, define the ordering  $\alpha \leq \beta$  for multi-indices  $\alpha, \beta \in \mathbb{Z}_{\geq 0}^k$  to mean that  $\alpha_i \leq \beta_i$  for each  $1 \leq i \leq k$ . Then we see that

$$(1.8) \quad \mathbf{F}(f) = \{\mathbf{x}^\alpha : \alpha \neq (0, \dots, 0), \alpha \leq \beta \text{ for some } \beta \in \Lambda(f)\}.$$

Note that we assume  $\mathbf{F}(f)$  contains only distinct elements.

Clearly, if we define

$$(1.9) \quad \mathbf{F}_{d,k} = \{\mathbf{x}^\alpha : 1 \leq |\alpha| \leq d\},$$

then for any polynomial  $f$  of degree  $d$ ,

$$(1.10) \quad \mathbf{F}(f) \subseteq \mathbf{F}_{d,k}.$$

In fact, typically  $\mathbf{F}(f)$  will be smaller than  $\mathbf{F}_{d,k}$ . We define  $R(f)$  to be the number of elements in  $\mathbf{F}(f)$ , and  $M(f)$  to be the sum of the total degrees of the elements in  $\mathbf{F}(f)$ . (We thus see that  $R(f), M(f)$  are analogous to  $R_{d,k}, M_{d,k}$ .) Finally, we define the multi-index  $\gamma(f) \in \mathbb{Z}_{\geq 0}^k$  to be the sum of all the multi-indices occurring in  $\mathbf{F}(f)$ .

**Example A.** For a simple example, if  $f$  is itself a monomial, say

$$f(\mathbf{x}) = \mathbf{x}^\delta,$$

for a fixed multi-index  $\delta = (d_1, \dots, d_k)$ , then in this case we would have

$$(1.11) \quad \mathbf{F}(f) = \{\mathbf{x}^\alpha : 0 \leq \alpha_i \leq d_i \text{ for } 1 \leq i \leq k, \alpha \neq (0, \dots, 0)\}.$$

Upon defining

$$\mathcal{D} = \prod_{1 \leq i \leq k} (d_i + 1),$$

a simple calculation shows that in this case the set  $\mathbf{F}(f)$  has cardinality

$$(1.12) \quad R(f) := \#\mathbf{F}(f) = \left( \sum_{0 \leq \alpha_1 \leq d_1} \cdots \sum_{0 \leq \alpha_k \leq d_k} 1 \right) - 1 = \mathcal{D} - 1,$$

and the sum of the total degrees of the multi-indices in the set  $\mathbf{F}(f)$  is

$$(1.13) \quad M(f) := \sum_{0 \leq \alpha_1 \leq d_1} \cdots \sum_{0 \leq \alpha_k \leq d_k} (\alpha_1 + \cdots + \alpha_k) = \frac{1}{2} \mathcal{D} (d_1 + \cdots + d_k).$$

We may also compute

$$(1.14) \quad \gamma(f) := \sum_{0 \leq \alpha_1 \leq d_1} \cdots \sum_{0 \leq \alpha_k \leq d_k} (\alpha_1, \dots, \alpha_k) = \frac{1}{2} \mathcal{D} (d_1, \dots, d_k).$$

Now we may state our second main result:

**Theorem 1.2.** *For any integer  $r \geq R(f)(d+1)$ , if  $q_i^{\frac{1}{2(r-M(f))}} < H_i < q_i^{\frac{1}{2} + \frac{1}{4(r-M(f))}}$  for each  $i = 1, \dots, k$ , then*

$$S_k(f; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{-r-M(f)+1}{4r(r-M(f))} + \varepsilon} (\mathbf{q}^{\gamma(f)})^{\frac{1}{4r(r-M(f))}} q_{\max}^{\frac{2rk}{4r(r-M(f))}} q_{\min}^{-\frac{M(f)}{4r(r-M(f))}},$$

*uniformly in  $\mathbf{N}$ , with implied constant dependent on  $r, d, k, \varepsilon$  and independent of the coefficients of  $f$ .*

**Corollary 1.2.1.** *Under the conditions of Theorem 1.2, if in addition  $q_1 = \dots = q_k = q$  for a fixed prime  $q$ ,*

$$S_k(f; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1-\frac{1}{r}} q^{\frac{k(r+1-M(f))}{4r(r-M(f))} + \varepsilon}.$$

These results rely on a multi-variable version of the Vinogradov Mean Value Theorem tailored to the set  $\mathbf{F}(f)$  (see Section 2.1). Note that the bounds on the right hand side are sharper than those of Theorem 1.1 and its corollary for any  $f$  such that the inclusion in (1.10) is strict, so that  $M(f) < M_{d,k}$ , in which case the bounds also hold for a larger range of  $r$  since  $R(f) < R_{d,k}$ .

**Example B.** We will highlight the strength of this second type of result by considering the particularly simple case of

$$(1.15) \quad f(\mathbf{x}) = x_1^{d_1} x_2^{d_2}$$

in dimension  $k = 2$  with fixed integers  $d_1 > d_2 \geq 1$ , and total degree  $d = d_1 + d_2$ . We compute that

$$\mathbf{F}(f) = \{x_1^{\alpha_1} x_2^{\alpha_2} : 0 \leq \alpha_1 \leq d_1, 0 \leq \alpha_2 \leq d_2, (\alpha_1, \alpha_2) \neq (0, 0)\}.$$

Thinking of  $d_2$  as fixed and  $d_1$  as arbitrarily large relative to  $d_2$ , we see from (1.12)–(1.14) that

$$\begin{aligned} R(f) &= (d_1 + 1)(d_2 + 1) - 1 \approx d_1 \\ M(f) &= \frac{1}{2}(d_1 + d_2)(d_1 + 1)(d_2 + 1) \approx d_1^2 \\ \gamma(f) &= \frac{1}{2}(d_1 + 1)(d_2 + 1)(d_1, d_2) \approx (d_1^2, d_1). \end{aligned}$$

In comparison, we see from (1.6) that

$$\begin{aligned} R_{d,k} &= \frac{1}{2}(d_1 + d_2 + 2)(d_1 + d_2 + 1) - 1 \approx d_1^2 \\ M_{d,k} &= \frac{1}{3}(d_1 + d_2 + 2)(d_1 + d_2 + 1)(d_1 + d_2) \approx d_1^3. \end{aligned}$$

Thus the bound provided by Corollary 1.2.1 is significantly sharper than that of Corollary 1.1.1, and the range  $r \geq R(f)(d+1)$  is longer than the range  $r \geq R_{d,k}(d+1)$ .

This is a genuinely multi-dimensional phenomenon. In the case of dimension  $k = 1$ , given any fixed polynomial  $f(x)$  of degree  $d$ , one necessarily computes

$$\mathbf{F}(f) = \{x, x^2, \dots, x^d\} = \mathbf{F}_{d,1};$$

that is, in dimension  $k = 1$ , equality always holds in (1.10). The strength of Theorem 1.2 stems from the fact that in the multi-variable setting, given a fixed polynomial  $f$ , the resulting set  $\mathbf{F}(f)$  is typically much smaller than  $\mathbf{F}_{d,k}$ .

**Example C.** In addition, we note that the explicit presence of the exponent  $\gamma(f)$  in Theorem 1.2 can also be advantageous, when the primes  $q_i$  have varying sizes. (Such a

situation can be encountered in applications, for example, which require counting integral points on a hypersurface within a box with disparate side-lengths.) Continuing with the example (1.15), moduli  $q_1, q_2$ , and degrees  $d_1, d_2$  with  $d_1$  arbitrarily large relative to  $d_2$ , the term  $\mathbf{q}^{\gamma(f)}$  in Theorem 1.2 takes the form (for some constants  $c_i$ )

$$\mathbf{q}^{\gamma(f)} \approx q_1^{c_1 d_1^2} q_2^{c_2 d_1};$$

this is advantageous compared to the analogous factor in Theorem 1.1, namely

$$\|\mathbf{q}\|^{\frac{M}{k}} \approx q_1^{c_3 d_1^3} q_2^{c_4 d_1^3},$$

if for example  $q_2$  is large compared to  $q_1$ . With these contrasting examples in mind, we now turn to the fully general setting in which we will work for the remainder of the paper.

## 2. THE GENERAL SETTING OF TRANSLATION-DILATION INVARIANT SYSTEMS

Let  $\mathbf{F}$  denote a system of homogeneous polynomials,

$$\mathbf{F} = \{F_1, \dots, F_R\}$$

with  $F_\ell \in \mathbb{Z}[X_1, \dots, X_k]$  for each  $1 \leq \ell \leq R$ . Consider for any integer  $r \geq 1$  the system of  $R$  simultaneous Diophantine equations

$$(2.1) \quad \sum_{j=1}^r (F_\ell(\mathbf{x}_j) - F_\ell(\mathbf{y}_j)) = \mathbf{0}, \quad \text{for all } 1 \leq \ell \leq R,$$

where  $\mathbf{x}_j, \mathbf{y}_j \in \mathbb{Z}^k$  for  $j = 1, \dots, r$ . Define  $J_r(\mathbf{F}; X)$  to be the number of integral solutions of the system (2.1) with  $1 \leq x_{j,i}, y_{j,i} \leq X$  for all  $1 \leq j \leq r, 1 \leq i \leq k$ . In [14], Parsell, Prendiville and Wooley prove strong upper bounds for  $J_r(\mathbf{F}; X)$  when  $\mathbf{F}$  is a translation-dilation invariant system, which we now define.

We say  $\mathbf{F}$  is a *translation-dilation invariant* system if the following two properties are satisfied: (i) the polynomials  $F_1, \dots, F_R$  are each homogeneous of positive degree; and (ii) there exist polynomials

$$c_{m\ell} \in \mathbb{Z}[\xi_1, \dots, \xi_k], \quad \text{for each } 1 \leq m \leq R, 0 \leq \ell \leq m,$$

with  $c_{mm} = 1$  for  $1 \leq m \leq R$ , such that for any  $\boldsymbol{\xi} \in \mathbb{Z}^k$ ,

$$(2.2) \quad F_m(\mathbf{x} + \boldsymbol{\xi}) = c_{m0}(\boldsymbol{\xi}) + \sum_{\ell=1}^m c_{m\ell}(\boldsymbol{\xi}) F_\ell(\mathbf{x}), \quad 1 \leq m \leq R.$$

(As in [14], we note that the number of solutions to (2.1) counted by  $J_r(\mathbf{F}; X)$  is not affected when one re-orders the  $F_\ell$  or takes independent linear combinations of the original forms; so we will say a system is translation-dilation invariant if it is equivalent via such manipulations to a system which is translation-dilation invariant in the strict sense.)

Translation-dilation invariant systems are simple to generate. As a first example, note that  $\mathbf{F}_{d,k}$  defined in (1.9) is a translation-dilation invariant system. As a second example, given any polynomial  $f$ , the set  $\mathbf{F}(f)$  constructed in (1.8) is a translation-dilation invariant system. In fact, more generally, given any collection of homogeneous polynomials, say

$$G_1, \dots, G_h \in \mathbb{Z}[X_1, \dots, X_k],$$

one can construct a translation-dilation invariant system. One first constructs the set  $\mathcal{G}$  consisting of all the partial derivatives

$$\frac{\partial^{t_1+\dots+t_k}}{\partial x_1^{t_1} \dots \partial x_k^{t_k}} G_m(\mathbf{x}), \quad 1 \leq m \leq h,$$

with integral  $t_i \geq 0$  for each  $1 \leq i \leq k$ . The set  $\mathcal{G}$  is clearly finite; let  $\mathcal{G}_0 = \{F_1, \dots, F_R\}$  denote the subset of  $\mathcal{G}$  consisting of all polynomials with positive degree, labeled so that  $\deg F_1 \leq \deg F_2 \leq \dots \leq \deg F_R$ . Then one confirms via the multi-dimensional Taylor's theorem that the conditions (2.2) hold, for some choice of coefficients  $c_{m\ell}(\boldsymbol{\xi}) \in \mathbb{Z}[\xi_1, \dots, \xi_k]$  such that  $c_{mm}(\boldsymbol{\xi}) = 1$  for  $1 \leq m \leq R$ . Furthermore, by replacing the set of forms  $\mathcal{G}_0$  by any subset whose span contains  $F_1, \dots, F_R$ , we may assume that the set  $\{F_1, \dots, F_R\}$  is linearly independent, in which case we say the system is *reduced*. Finally, we introduce the notion of a *monomial* translation-dilation invariant system, simply by requiring that each form  $F_\ell$  in the system be a monomial. We will also avoid certain vacuous cases by making explicit the requirement that a reduced monomial translation-dilation invariant system of dimension  $k$  in variables  $X_1, \dots, X_k$  includes for each  $i = 1, \dots, k$  at least one monomial of positive degree with respect to  $X_i$ . To summarize, we may conclude that for any polynomial  $f$  we will consider, the set  $\mathbf{F}(f)$  is a reduced monomial translation-dilation invariant system.

We now define the parameters used in [14] to characterize a reduced monomial translation-dilation invariant system  $\mathbf{F} = \{F_1, \dots, F_R\}$  with monomials  $F_\ell \in \mathbb{Z}[X_1, \dots, X_k]$ . We say that  $k = k(\mathbf{F})$  is the *dimension* of the system and  $R = R(\mathbf{F})$  is the *rank*. For each monomial  $F_\ell$  we let  $d_\ell(\mathbf{F}) = \deg(F_\ell)$  be the total degree of the monomial. We define the *degree*  $d = d(\mathbf{F})$  of the system by

$$d(\mathbf{F}) = \max_{1 \leq \ell \leq R} d_\ell(\mathbf{F}).$$

We define the *weight*  $M = M(\mathbf{F})$  of the system by

$$M(\mathbf{F}) = \sum_{\ell=1}^R d_\ell(\mathbf{F}).$$

It is also convenient to use an alternative representation of  $\mathbf{F} = \{F_1, \dots, F_R\}$  by explicitly writing  $\mathbf{F}$  as a collection of monomials

$$\{\mathbf{x}^\beta : \beta \in \Lambda(\mathbf{F})\},$$

for a fixed collection  $\Lambda(\mathbf{F})$  of  $R$  distinct non-zero multi-indices  $\beta \in \mathbb{Z}_{\geq 0}^k$ . If  $\mathbf{F}$  has degree  $d$ , then we see that  $|\beta| \leq d$  for each  $\beta \in \Lambda(\mathbf{F})$  (and there exists some  $\beta \in \Lambda(\mathbf{F})$  with  $|\beta| = d$ ), and the rank  $R(\mathbf{F})$  is  $|\Lambda(\mathbf{F})|$ . The weight is

$$M(\mathbf{F}) = \sum_{\beta \in \Lambda(\mathbf{F})} |\beta|.$$

Finally, we define the notion of the *density*  $\gamma = \gamma(\mathbf{F}) \in \mathbb{Z}_{\geq 0}^k$  of the system by setting

$$(2.3) \quad \gamma(\mathbf{F}) = \sum_{\beta \in \Lambda(\mathbf{F})} \beta.$$

In particular, we note that  $|\gamma| = M(\mathbf{F})$ .

**2.1. Vinogradov Mean Value Theorem.** We recall the main result of Parsell, Pridmore and Wooley in full generality:

**Theorem B** (Theorem 2.1 of [14]). *Let  $\mathbf{F}$  be a reduced translation-dilation invariant system having dimension  $k$ , degree  $d$ , rank  $R$  and weight  $M$ . Suppose that  $r$  is a natural number with  $r \geq R(d+1)$ . Then for each  $\varepsilon > 0$ ,*

$$(2.4) \quad J_r(\mathbf{F}; X) \ll X^{2rk-M+\varepsilon}.$$

Theorem A corresponds to the special case of taking  $\mathbf{F}$  to be the system  $\mathbf{F}_{d,k}$  in (1.9).

**2.2. Statement of general results.** Our main result in full generality is:

**Theorem 2.1.** *Let  $\mathbf{F}$  be a reduced monomial translation-dilation invariant system having dimension  $k$ , degree  $d$ , rank  $R$ , weight  $M$ , and density  $\gamma$ . Let  $\mathcal{F}$  denote the set of all real-valued polynomials spanned by the system  $\mathbf{F}$ . If  $r > R(d+1)$  and  $q_i^{\frac{1}{2(r-M)}} < H_i < q_i^{\frac{1}{2} + \frac{1}{4(r-M)}}$  for each  $i = 1, \dots, k$ , then*

$$\sup_{f \in \mathcal{F}} |S_k(f; \mathbf{N}, \mathbf{H})| \ll_{r,d,k,\varepsilon} \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{-r-M+1}{4r(r-M)}+\varepsilon} (\mathbf{q}^\gamma)^{\frac{1}{4r(r-M)}} q_{\max}^{\frac{2rk}{4r(r-M)}} q_{\min}^{-\frac{M}{4r(r-M)}},$$

uniformly in  $\mathbf{N}$ .

**Corollary 2.1.1.** *Under the hypotheses of Theorem 2.1, if in addition  $q_1 = \dots = q_k = q$  for a fixed prime  $q$ ,*

$$\sup_{f \in \mathcal{F}} |S_k(f; \mathbf{N}, \mathbf{H})| \ll_{r,d,k,\varepsilon} \|\mathbf{H}\|^{1-\frac{1}{r}} q^{\frac{k(r+1-M)}{4r(r-M)}+\varepsilon}.$$

As usual, we may check the strength of this result by computing that if  $H_i = q^{1/4+\kappa_i}$  for each  $i = 1, \dots, k$  then Corollary 2.1.1 provides a bound of size  $\|\mathbf{H}\|q^{-\delta}$  where

$$(2.5) \quad \delta \approx \frac{\left(\sum_{i=1}^k \kappa_i\right)^2}{k}.$$

(See Section 6.2 for details.) Notably, this is independent of the degree, rank, and weight of the system  $\mathbf{F}$ , and only dependent on the dimension  $k$ . This also recovers the strength of the original Burgess bound (1.3) in dimension  $k = 1$ .

We note that the input of Theorem B is crucial; if we used the Burgess method alone without inputting an appropriate Vinogradov Mean Value Theorem, we would obtain a result with

$$(2.6) \quad \delta \approx \frac{\left(\sum_{i=1}^k \kappa_i\right)^2}{M+k}$$

in place of (2.5), which is weaker both because it is smaller and because it is dependent on the degree  $d$  of the polynomial  $f$ . (We will record such a result later in Theorem 4.6.) Finally, it is clear that Theorems 1.1 and 1.2 are immediate corollaries of Theorem 2.1.

We remark that the approach of this paper is expected to generalize, when suitably adapted, to translation-dilation invariant systems of homogeneous polynomials that are not necessarily monomials. Additionally, we note that  $\{x_1, \dots, x_k\}$  is a special case of a system



of  $k$  linearly independent linear forms over  $\mathbb{F}_q$ . In [6] Burgess proved that if  $\{L_i\}_{1 \leq i \leq k}$  is a system of  $k$  linearly independent linear forms over  $\mathbb{F}_q$  for  $q$  prime, then

$$(2.7) \quad \sum_{\substack{\mathbf{n} \in \mathbb{Z}^k \\ \mathbf{n} \in (\mathbf{N}, \mathbf{N} + \mathbf{H}]}} \chi\left(\prod_{i=1}^k L_i(\mathbf{n})\right) \ll H^k q^{-\delta}$$

for some small  $\delta = \delta(k) > 0$ , provided  $H > q^{\frac{1}{2} - \frac{1}{2k+2} + \varepsilon}$ . More recently, Chang [8] (for  $k = 2$ ) and Bourgain and Chang [2] (for  $k \geq 3$ ) have proved a bound of the form (2.7) that is nontrivial in the original Burgess range of  $H > q^{1/4 + \varepsilon}$ . It is reasonable to expect that the methods of this paper will generalize to mixed character sums involving products of linear forms of this type.

**2.3. Notation.** For two  $k$ -tuples  $\mathbf{K} = (K_1, \dots, K_k)$  and  $\mathbf{H} = (H_1, \dots, H_k)$  of real numbers, we will let  $\mathbf{K} \leq \mathbf{H}$  represent that all the following conditions hold:

$$K_1 \leq H_1, \dots, K_k \leq H_k.$$

We define  $\mathbf{K} < \mathbf{H}$  and  $\mathbf{K} \ll \mathbf{H}$  similarly. We will denote by  $\mathbf{K} \circ \mathbf{H}$  the coordinate-wise product,

$$\mathbf{K} \circ \mathbf{H} = (K_1 H_1, \dots, K_k H_k).$$

We will write

$$\mathbf{K}^{-1} = (K_1^{-1}, \dots, K_k^{-1}),$$

and use the notation

$$\mathbf{H}/\mathbf{K} = \mathbf{H} \circ \mathbf{K}^{-1} = (H_1/K_1, \dots, H_k/K_k).$$

For any  $k$ -tuple  $\mathbf{K} = (K_1, \dots, K_k)$  we set

$$\|\mathbf{K}\| = K_1 \cdots K_k.$$

For a scalar  $q$ , we will say that  $\mathbf{K} = (K_1, \dots, K_k)$  is regarded modulo  $q$  if each  $K_i$  is regarded modulo  $q$ . We will say  $\mathbf{K}$  is regarded modulo  $\mathbf{H}$  if  $K_i$  is regarded modulo  $H_i$  for each  $i = 1, \dots, k$ . For a scalar  $q$ , we will write  $\mathbf{K}q = (K_1 q, \dots, K_k q)$ . We will let implied constants depend on  $r, d, k$  and  $\varepsilon$  as appropriate. We define the notation  $\mathcal{L}(\mathbf{q}) = \prod \log q_i$ .

### 3. ACTIVATION OF THE BURGESS METHOD

Let  $\mathbf{F} = \{F_1, \dots, F_R\}$  be a given reduced monomial translation-dilation invariant system of dimension  $k$ , degree  $d$ , rank  $R$ , weight  $M$  and density  $\gamma$ . We will let  $\Lambda(\mathbf{F})$  be the associated set of multi-indices, so that we can represent  $\mathbf{F}$  as  $\{\mathbf{x}^\beta : \beta \in \Lambda(\mathbf{F})\}$ . We will let  $\mathcal{F}(\mathbf{F})$  denote the set of all real-valued polynomials spanned by the set of monomials comprising  $\mathbf{F}$ . We will let  $\mathcal{F}_0(\mathbf{F})$  denote the set of all real-valued polynomials spanned by  $1 \cup \mathbf{F}$ ; that is, we expand  $\mathcal{F}(\mathbf{F})$  to include polynomials with constant terms. We correspondingly set  $\Lambda_0(\mathbf{F}) = \{(0, \dots, 0)\} \cup \Lambda(\mathbf{F})$ .

The family  $\mathcal{F}_0(\mathbf{F})$  is invariant under translations: by the relations (2.2), if  $f(\mathbf{x}) \in \mathcal{F}_0(\mathbf{F})$  then  $f(\mathbf{x} + \boldsymbol{\xi}) \in \mathcal{F}_0(\mathbf{F})$  for all  $\boldsymbol{\xi} \in \mathbb{R}^k$ . Similarly  $\mathcal{F}_0(\mathbf{F})$  is invariant under dilations  $\mathbf{x} \mapsto \boldsymbol{\xi} \circ \mathbf{x}$ : that is, if  $f(\mathbf{x}) \in \mathcal{F}_0(\mathbf{F})$  then  $f(\boldsymbol{\xi} \circ \mathbf{x}) \in \mathcal{F}_0(\mathbf{F})$  for all  $\boldsymbol{\xi} \in \mathbb{R}^k$ . This is a stronger type of dilation invariance than dilation by scalars, and is a consequence of using monomial systems. To confirm this, we simply represent  $f$  as

$$f(\mathbf{x}) = \sum_{\beta \in \Lambda_0(\mathbf{F})} f_\beta \mathbf{x}^\beta$$

with coefficients  $f_\beta$ , so that

$$f(\boldsymbol{\xi} \circ \mathbf{x}) = \sum_{\beta \in \Lambda_0(\mathbf{F})} f_\beta(\boldsymbol{\xi} \circ \mathbf{x})^\beta = \sum_{\beta \in \Lambda_0(\mathbf{F})} (f_\beta \boldsymbol{\xi}^\beta) \mathbf{x}^\beta,$$

which is also a polynomial in  $\mathcal{F}_0(\mathbf{F})$ . Finally, we note that since we assume in the definition of a reduced monomial translation-dilation invariant system that for each  $i = 1, \dots, k$ ,  $\mathbf{F}$  contains a monomial of positive degree in  $X_i$ , expanding the relations (2.2) using the multinomial theorem shows that linear monomials in each of  $X_1, \dots, X_k$  also belong to  $\mathbf{F}$ . We will use these facts repeatedly in the argument to come.

From now on  $\mathbf{F}$  will be the fixed system given above. Fix primes  $q_1, \dots, q_k$  (not necessarily distinct) and let  $\mathbf{q} = (q_1, \dots, q_k)$ . For each  $i = 1, \dots, k$  let  $\chi_i$  be a non-principal multiplicative Dirichlet character modulo  $q_i$ . Instead of working directly with  $S_k(f; \mathbf{N}, \mathbf{H})$  we will define

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) = \sup_{f \in \mathcal{F}_0(\mathbf{F})} \sup_{\mathbf{K} \leq \mathbf{H}} \left| \sum_{\substack{\mathbf{x} \in \mathbb{Z}^k \\ \mathbf{x} \in (\mathbf{N}, \mathbf{N} + \mathbf{K}]}} e(f(\mathbf{x})) \chi_1(x_1) \cdots \chi_k(x_k) \right|,$$

which certainly majorizes  $S_k(f; \mathbf{N}, \mathbf{H})$ . We first note that  $T(\mathbf{F}; \mathbf{N}, \mathbf{H})$  is unchanged if the supremum over  $f \in \mathcal{F}_0(\mathbf{F})$  is restricted to  $f \in \mathcal{F}(\mathbf{F})$ , as appears in the statement of our theorems. Second, we note that  $T(\mathbf{F}; \mathbf{N}, \mathbf{H})$  is periodic modulo  $\mathbf{q}$  with respect to  $\mathbf{N}$ . Indeed, if  $\mathbf{N} = \mathbf{M} \circ \mathbf{q} + \mathbf{L}$  for an integer tuple  $\mathbf{M}$ , we can express  $T(\mathbf{F}; \mathbf{N}, \mathbf{H})$  as

$$\begin{aligned} & \sup_{f \in \mathcal{F}_0(\mathbf{F})} \sup_{\mathbf{K} \leq \mathbf{H}} \left| \sum_{\substack{\mathbf{x} \in \mathbb{Z}^k \\ \mathbf{x} \in (\mathbf{L}, \mathbf{L} + \mathbf{K}]}} e(f(\mathbf{x} + \mathbf{M} \circ \mathbf{q})) \chi_1(x_1 + M_1 q_1) \cdots \chi_k(x_k + M_k q_k) \right| \\ &= \sup_{f \in \mathcal{F}_0(\mathbf{F})} \sup_{\mathbf{K} \leq \mathbf{H}} \left| \sum_{\substack{\mathbf{x} \in \mathbb{Z}^k \\ \mathbf{x} \in (\mathbf{L}, \mathbf{L} + \mathbf{K}]}} e(f(\mathbf{x} + \mathbf{M} \circ \mathbf{q})) \chi_1(x_1) \cdots \chi_k(x_k) \right|, \end{aligned}$$

which is  $T(\mathbf{F}; \mathbf{L}, \mathbf{H})$ , as claimed. Thus we see that it suffices to consider  $\mathbf{N}$  with  $0 \leq N_i < q_i$  for  $i = 1, \dots, k$ . We also note that in  $T(\mathbf{F}; \mathbf{N}, \mathbf{H})$  it suffices to regard the coefficients of the polynomial  $f$  modulo 1; by a compactness argument, one sees that the value of  $T(\mathbf{F}; \mathbf{N}, \mathbf{H})$  is achieved by a particular choice of polynomial  $f$  and length  $\mathbf{K}$ .

We now begin the familiar opening gambit of the Burgess method. Given a fixed  $\mathbf{H} = (H_1, \dots, H_k)$ , we let  $P_1, \dots, P_k$  be a set of parameters each satisfying  $1 \leq P_i \leq H_i$ , to be chosen precisely later. For each  $i = 1, \dots, k$  we fix a set of primes

$$\mathcal{P}_i = \{P_i < p \leq 2P_i\}.$$

We then let  $\mathcal{P}$  denote the corresponding set of  $k$ -tuples of primes:

$$\mathcal{P} = \{\mathbf{p} = (p_1, \dots, p_k) : p_i \in \mathcal{P}_i, i = 1, \dots, k\}.$$

Since we will restrict to  $H_i = o(q_i)$  in our theorems, we will be able to assume  $p_i \nmid q_i$  for all  $p_i \in \mathcal{P}_i$ , for all  $i$ . We also note that for each  $i$ ,  $|\mathcal{P}_i| \gg P_i (\log P_i)^{-1} \gg P_i (\log q_i)^{-1}$ , so that

$$(3.1) \quad |\mathcal{P}| \gg P_1 \cdots P_k \left( \prod_{i=1}^k \log q_i \right)^{-1} = \|\mathbf{P}\| \mathcal{L}(\mathbf{q})^{-1}.$$

Fix a tuple  $\mathbf{K} \leq \mathbf{H}$  and a tuple  $\mathbf{p}$  of primes in  $\mathcal{P}$ ; then each  $\mathbf{x} \in (\mathbf{N}, \mathbf{N} + \mathbf{K}]$  may be split into residue classes modulo  $\mathbf{p}$ , so that for each  $i = 1, \dots, k$ , we may write

$$x_i = a_i q_i + p_i m_i,$$

where  $0 \leq a_i < p_i$  and  $m_i \in (N_i^{a_i, p_i}, N_i^{a_i, p_i} + K_i^{a_i, p_i}]$ , where we have set

$$\begin{aligned} N_i^{a_i, p_i} &= \frac{N_i - a_i q_i}{p_i}, \\ K_i^{a_i, p_i} &= \frac{K_i}{p_i} \leq \frac{H_i}{p_i} \leq \frac{H_i}{P_i}. \end{aligned}$$

That is to say,  $\mathbf{x} = \mathbf{a} \circ \mathbf{q} + \mathbf{p} \circ \mathbf{m}$  with  $\mathbf{0} \leq \mathbf{a} < \mathbf{p}$  and  $\mathbf{m} \in [\mathbf{N}^{\mathbf{a}, \mathbf{p}}, \mathbf{N}^{\mathbf{a}, \mathbf{p}} + \mathbf{K}^{\mathbf{a}, \mathbf{p}}]$ . Then we see that

$$\begin{aligned} (3.2) \quad & \sum_{\substack{\mathbf{x} \in \mathbb{Z}^k \\ \mathbf{x} \in (\mathbf{N}, \mathbf{N} + \mathbf{K}]}} e(f(\mathbf{x})) \chi_1(x_1) \cdots \chi_k(x_k) \\ &= \sum_{\mathbf{0} \leq \mathbf{a} < \mathbf{p}} \sum_{\mathbf{m} \in (\mathbf{N}^{\mathbf{a}, \mathbf{p}}, \mathbf{N}^{\mathbf{a}, \mathbf{p}} + \mathbf{K}^{\mathbf{a}, \mathbf{p}}]} e(f(\mathbf{a} \circ \mathbf{q} + \mathbf{p} \circ \mathbf{m})) \prod_{i=1}^k \chi_i(a_i q_i + p_i m_i). \end{aligned}$$

We may remove the dependence of the multiplicative characters on  $p_i$ , since

$$\prod_{i=1}^k \chi_i(a_i q_i + p_i m_i) = \prod_{i=1}^k \chi_i(m_i) \cdot \prod_{i=1}^k \chi_i(p_i).$$

Thus after taking absolute values and taking the supremum over  $f \in \mathcal{F}_0(\mathbf{F})$  and  $\mathbf{K} \leq \mathbf{H}$  in (3.2), we see that

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \leq \sum_{\mathbf{0} \leq \mathbf{a} < \mathbf{p}} T(\mathbf{F}; \mathbf{N}^{\mathbf{a}, \mathbf{p}}, \mathbf{H}/\mathbf{P}).$$

After averaging over the set  $\mathcal{P}$ , we then have

$$(3.3) \quad T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \leq |\mathcal{P}|^{-1} \sum_{\mathbf{p} \in \mathcal{P}} \sum_{\mathbf{0} \leq \mathbf{a} < \mathbf{p}} T(\mathbf{F}; \mathbf{N}^{\mathbf{a}, \mathbf{p}}, \mathbf{H}/\mathbf{P}).$$

We will now make the starting points  $\mathbf{N}^{\mathbf{a}, \mathbf{p}}$  of the sums  $T(\mathbf{F}; \mathbf{N}^{\mathbf{a}, \mathbf{p}}, \mathbf{H}/\mathbf{P})$  independent of  $\mathbf{a}, \mathbf{p}$  via the following lemma:

**Lemma 3.1.** *For any tuple  $\mathbf{U}$  of real numbers and tuple  $\mathbf{L}$  of real numbers with  $L_i \geq 1$  for  $i = 1, \dots, k$ ,*

$$T(\mathbf{F}; \mathbf{U}, \mathbf{L}) \leq 2^{2k} \|\mathbf{L}\|^{-1} \sum_{\mathbf{U} - \mathbf{L} < \mathbf{m} \leq \mathbf{U}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{L}).$$

Suppose  $T(\mathbf{F}; \mathbf{U}, \mathbf{L})$  is attained by a certain polynomial  $f$  and a tuple  $\mathbf{K} \leq \mathbf{L}$ ; then we write

$$T(\mathbf{F}; \mathbf{U}, \mathbf{L}) = \left| \sum_{\mathbf{x} \in (\mathbf{U}, \mathbf{U} + \mathbf{K}]} e(f(\mathbf{x})) \prod_{i=1}^k \chi_i(x_i) \right|.$$

By the inclusion-exclusion principle, for any fixed  $\mathbf{R}$  with  $\mathbf{U} - \mathbf{L} < \mathbf{R} \leq \mathbf{U}$ ,

$$\sum_{\mathbf{x} \in (\mathbf{U}, \mathbf{U} + \mathbf{K}]} e(f(\mathbf{x})) \prod_{i=1}^k \chi_i(x_i) = \sum_{\substack{\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_k) \\ \varepsilon_i \in \{0, 1\}}} (-1)^{|\boldsymbol{\varepsilon}|} \sum_{\mathbf{x} \in (\mathbf{R}, (\mathbf{1} - \boldsymbol{\varepsilon}) \circ \mathbf{K} + \mathbf{U}]} e(f(\mathbf{x})) \prod_{i=1}^k \chi_i(x_i).$$

Here  $|\varepsilon| = \varepsilon_1 + \dots + \varepsilon_k$ , and  $\mathbf{1} - \varepsilon = (1 - \varepsilon_1, \dots, 1 - \varepsilon_k)$ .

We next note that for any  $\mathbf{R}$  with  $\mathbf{U} - \mathbf{L} < \mathbf{R} \leq \mathbf{U}$  and any  $\varepsilon$  as above, the side-lengths of the box  $(\mathbf{R}, (\mathbf{1} - \varepsilon) \circ \mathbf{K} + \mathbf{U}]$  satisfy

$$(\mathbf{1} - \varepsilon) \circ \mathbf{K} + \mathbf{U} - \mathbf{R} \leq 2\mathbf{L}.$$

Thus

$$\left| \sum_{\mathbf{x} \in (\mathbf{U}, \mathbf{U} + \mathbf{K}]} e(f(\mathbf{x})) \prod_{i=1}^k \chi_i(x_i) \right| \leq 2^k T(\mathbf{F}; \mathbf{R}, 2\mathbf{L}).$$

We finally note that there are at least  $L_i/2$  integers in the interval  $(U_i - L_i, U_i]$  and hence at least  $2^{-k} \|\mathbf{L}\|$  choices for tuples  $\mathbf{R}$  in the box  $(\mathbf{U} - \mathbf{L}, \mathbf{U}]$ , so that averaging over all of these choices produces the result of Lemma 3.1.

We now apply Lemma 3.1 to (3.3) with the choice  $\mathbf{L} = \mathbf{H}/\mathbf{P}$ , to see by (3.1) that

$$\begin{aligned} T(\mathbf{F}; \mathbf{N}, \mathbf{H}) &\leq 2^{2k} \|\mathbf{H}/\mathbf{P}\|^{-1} |\mathcal{P}|^{-1} \sum_{\mathbf{p} \in \mathcal{P}} \sum_{0 \leq \mathbf{a} < \mathbf{p}} \sum_{\mathbf{N}^{\mathbf{a}, \mathbf{p}} - \mathbf{H}/\mathbf{P} < \mathbf{m} \leq \mathbf{N}^{\mathbf{a}, \mathbf{p}}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P}) \\ &\ll \|\mathbf{H}\|^{-1} \mathcal{L}(\mathbf{q}) \sum_{\mathbf{p} \in \mathcal{P}} \sum_{0 \leq \mathbf{a} < \mathbf{p}} \sum_{\mathbf{N}^{\mathbf{a}, \mathbf{p}} - \mathbf{H}/\mathbf{P} < \mathbf{m} \leq \mathbf{N}^{\mathbf{a}, \mathbf{p}}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P}). \end{aligned}$$

Now for each  $\mathbf{m}$  we define  $\mathcal{A}(\mathbf{m})$  to be the quantity

$$\#\{\mathbf{a}, \mathbf{p} : 0 \leq a_i < p_i \text{ and } p_i \in \mathcal{P}_i : \frac{N_i - a_i q_i}{p_i} - \frac{H_i}{P_i} < m_i \leq \frac{N_i - a_i q_i}{p_i}, i = 1, \dots, k\}.$$

With this notation, we may now write

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{-1} \mathcal{L}(\mathbf{q}) \sum_{\mathbf{m} \in \mathbb{Z}^k} \mathcal{A}(\mathbf{m}) T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P}).$$

We now define

$$S_1 = \sum_{\mathbf{m}} \mathcal{A}(\mathbf{m}),$$

and

$$S_2 = \sum_{\mathbf{m}} \mathcal{A}(\mathbf{m})^2.$$

We record the following facts, which we prove in Section 7.1:

**Lemma 3.2.** *We have  $\mathcal{A}(\mathbf{m}) = 0$  unless  $|m_i| \leq 2q_i$  for  $i = 1, \dots, k$ . Furthermore if  $H_i P_i < q_i$  for each  $i = 1, \dots, k$  then*

$$S_1 \leq S_2 \ll \|\mathbf{H}\| \|\mathbf{P}\|.$$

After a repeated application of Hölder's inequality, Lemma 3.2 allows us to conclude that

$$\begin{aligned} T(\mathbf{F}; \mathbf{N}, \mathbf{H}) &\ll \|\mathbf{H}\|^{-1} \mathcal{L}(\mathbf{q}) S_1^{1-\frac{1}{r}} S_2^{\frac{1}{2r}} \left\{ \sum_{\mathbf{m}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P})^{2r} \right\}^{\frac{1}{2r}} \\ &\ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \mathcal{L}(\mathbf{q}) \left\{ \sum_{\mathbf{m}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P})^{2r} \right\}^{\frac{1}{2r}}. \end{aligned}$$

We now recall that  $T(\mathbf{F}; \mathbf{m}, \mathbf{K})$  is periodic in  $\mathbf{m}$  with respect to  $\mathbf{q}$ , so that it suffices to write

$$(3.4) \quad T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \mathcal{L}(\mathbf{q}) \left\{ \sum_{\mathbf{m} \pmod{\mathbf{q}}} T(\mathbf{F}; \mathbf{m}, 2\mathbf{H}/\mathbf{P})^{2r} \right\}^{\frac{1}{2r}}.$$

We now make the step of removing the supremum over lengths in the definition of  $T(\mathbf{F}; \mathbf{N}, \mathbf{K})$ . We define for any tuples  $\mathbf{M}, \mathbf{K}$  with  $K_i > 0$  the sum

$$T_0(\mathbf{F}; \mathbf{M}, \mathbf{K}) = \sup_{f \in \mathcal{F}_0(\mathbf{F})} \left| \sum_{\mathbf{M} < \mathbf{x} \leq \mathbf{M} + \mathbf{K}} e(f(\mathbf{x})) \chi_1(x_1) \cdots \chi_k(x_k) \right|.$$

We will use the following lemma, a  $k$ -dimensional version of Lemma 2.2 of Bombieri and Iwaniec [1], whose proof we indicate in Section 7.2.

**Lemma 3.3.** *Let  $a(\mathbf{n})$  be a sequence of complex numbers indexed by tuples  $\mathbf{n}$  supported on the set  $\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}] \subset \mathbb{Z}^k$ . Let  $I = (\mathbf{C}, \mathbf{C} + \mathbf{D}]$  be any product of intervals with  $I \subseteq (\mathbf{A}, \mathbf{A} + \mathbf{B}]$ . Then*

$$\sum_{\mathbf{n} \in I} a(\mathbf{n}) \ll \left( \prod_{i=1}^k \log(B_i + 2) \right) \sup_{\theta \in \mathbb{R}^k} \left| \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n}) e(\theta \cdot \mathbf{n}) \right|.$$

This lemma allows us to relate  $T(\mathbf{F}; \mathbf{M}, \mathbf{K})$  to  $T_0(\mathbf{F}; \mathbf{M}, \mathbf{K})$  since as long as  $d \geq 1$ , Lemma 3.3 shows that

$$T(\mathbf{F}; \mathbf{M}, \mathbf{K}) \ll \left( \prod_{i=1}^k \log(K_i + 2) \right) T_0(\mathbf{F}; \mathbf{M}, \mathbf{K}).$$

Note that here we use the assumption that  $d \geq 1$ , so that the linear exponential factor accrued in the application of Lemma 3.3 is absorbed in the supremum over polynomials  $f \in \mathcal{F}_0(\mathbf{F})$ . We also henceforward assume that  $K_i < q_i$  for each  $i = 1, \dots, k$ , so that the logarithmic factor is bounded above by  $\ll \mathcal{L}(\mathbf{q})$ ; this condition will be satisfied by our final choice of  $K_i$ , as we will later verify.

We may now re-write (3.4) as

$$(3.5) \quad T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \mathcal{L}(\mathbf{q})^2 S_3(2\mathbf{H}/\mathbf{P})^{\frac{1}{2r}},$$

where we define

$$(3.6) \quad S_3(\mathbf{K}) := \sum_{\mathbf{m} \pmod{\mathbf{q}}} T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r}.$$

#### 4. APPROXIMATION OF POLYNOMIALS

We will now bound  $S_3(\mathbf{K})$ , focusing first on an individual sum  $T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})$ ; recall that we assume from now on that each  $K_i < q_i$ . As in [13], the key step is to remove the supremum over all polynomials in  $\mathcal{F}_0(\mathbf{F})$  by showing, roughly speaking, that two polynomials with coefficients that are sufficiently close may be regarded as producing equivalent contributions, and thus we will majorize the supremum by summing over a collection of representative polynomials. We first perform a dissection of the coefficient space of  $\mathcal{F}_0(\mathbf{F})$ , recalling that we may regard the coefficients of any  $f \in \mathcal{F}_0(\mathbf{F})$  modulo 1.

We recall the collection of multi-indices  $\Lambda_0(\mathbf{F}) = \{(0, \dots, 0)\} \cup \Lambda(\mathbf{F})$  associated to the system  $\mathbf{F}$ . Since  $\mathbf{F}$  has rank  $R$ , we have  $R = |\Lambda(\mathbf{F})|$  and  $R + 1 = |\Lambda_0(\mathbf{F})|$ , so that  $R + 1$  is the dimension of the coefficient space of  $\mathcal{F}_0(\mathbf{F})$ .

Fix positive integers  $Q_1, \dots, Q_k$  and set  $\mathbf{Q} = (Q_1, \dots, Q_k)$ . We will choose  $Q_i$  precisely later; for now we assume that  $Q_i \geq K_i$  for each  $i$ , which we will verify later. We index the coefficient space  $[0, 1]^{R+1}$  as

$$[0, 1]^{R+1} = [0, 1] \times \dots \times [0, 1] = \prod_{\beta \in \Lambda_0(\mathbf{F})} [0, 1]^{(\beta)}.$$

For each of the  $R + 1$  multi-indices  $\beta \in \Lambda_0(\mathbf{F})$ , we partition the corresponding unit interval  $[0, 1]^{(\beta)} = [0, 1]$  indexed by  $\beta$  into  $\mathbf{Q}^\beta = Q_1^{\beta_1} \dots Q_k^{\beta_k}$  sub-intervals of length  $(\mathbf{Q}^\beta)^{-1}$ . We claim this partitions the full space  $[0, 1]^{R+1}$  into  $\mathbf{Q}^\gamma$  boxes, where we recall that  $\gamma = \gamma(\mathbf{F})$  is the density of the system  $\mathbf{F}$ , as defined in (2.3). We may verify this as follows: clearly the number of boxes is

$$\prod_{\beta \in \Lambda_0(\mathbf{F})} \mathbf{Q}^\beta = \mathbf{Q}^\delta,$$

say, where we have defined

$$\delta = \sum_{\beta \in \Lambda_0(\mathbf{F})} \beta = \sum_{\beta \in \Lambda(\mathbf{F})} \beta.$$

This last expression is precisely the definition of the density  $\gamma = \gamma(\mathbf{F})$ .

We will denote this dissection of the coefficient space as a union

$$(4.1) \quad [0, 1]^{R+1} = \bigcup_{\alpha} B_{\alpha}$$

over  $\mathbf{Q}^\gamma$  many boxes  $B_{\alpha}$ ; we may think of  $\alpha$  as a parameter in  $\mathbb{Z}_{\geq 0}$  indexing over a fixed ordering of the boxes. We will also associate to each box  $B_{\alpha}$  the fixed tuple  $\theta_{\alpha} \in B_{\alpha}$  that is the vertex of  $B_{\alpha}$  with the least value in each coordinate. Thus if we have fixed some enumeration  $\beta^{(0)}, \dots, \beta^{(R)}$  of the  $R + 1$  distinct multi-indices  $\beta \in \Lambda_0(\mathbf{F})$ , the distinguished vertex  $\theta_{\alpha}$  of a box  $B_{\alpha}$  takes the form

$$(4.2) \quad \theta_{\alpha} = (\theta_{\alpha, \beta^{(0)}}, \dots, \theta_{\alpha, \beta^{(R)}}) = (c_{\beta^{(0)}} \mathbf{Q}^{-\beta^{(0)}}, \dots, c_{\beta^{(R)}} \mathbf{Q}^{-\beta^{(R)}}),$$

where for each  $j = 0, \dots, R$ ,  $c_{\beta_j}$  is an integer with  $0 \leq c_{\beta_j} \leq \mathbf{Q}^{\beta_j} - 1$ . Finally, for any fixed  $\theta \in [0, 1]^{R+1}$ , we define an associated real-valued polynomial on  $\mathbb{R}^k$  by

$$(4.3) \quad \theta(\mathbf{X}) := \sum_{\beta \in \Lambda_0(\mathbf{F})} \theta_{\beta} \mathbf{X}^{\beta}.$$

We note that for any  $\theta \in [0, 1]^{R+1}$  this polynomial belongs to  $\mathcal{F}_0(\mathbf{F})$ .

For any tuple  $\mathbf{m}$  of integers and any tuple  $\boldsymbol{\tau}$  of positive real numbers and a fixed index  $\alpha$  of a box  $B_{\alpha}$  with associated vertex  $\theta_{\alpha}$ , we define

$$(4.4) \quad T(\alpha, \mathbf{F}; \mathbf{m}, \boldsymbol{\tau}) := \left| \sum_{\mathbf{0} \leq \mathbf{n} \leq \boldsymbol{\tau}} e(\theta_{\alpha}(\mathbf{n})) \chi_1(n_1 + m_1) \dots \chi_k(n_k + m_k) \right|.$$

Roughly speaking, our goal is to show that for any  $\mathbf{m}, \mathbf{K}$  there exists a suitable  $\alpha$  such that  $T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})$  (which takes a supremum over  $f \in \mathcal{F}_0(\mathbf{F})$ ) is well approximated by  $T(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K})$  (which corresponds to the single polynomial  $\theta_{\alpha}(\mathbf{X})$ ). In order to do so, we must use summation by parts, for which we require some notation.

Given any partition  $I \cup J$  of the set of indices  $\{1, \dots, k\}$  and a  $k$ -tuple  $\mathbf{n}$ , we will let  $\mathbf{n}_{(I)}$  denote the tuple of  $n_j$  with  $j \in I$  and similarly  $\mathbf{n}_{(J)}$  the tuple of  $n_j$  with  $j \in J$ ; thus for example we may write  $\mathbf{n} = (\mathbf{n}_{(I)}, \mathbf{n}_{(J)})$  (with some abuse of notation with respect to ordering). Given a sequence  $a(\mathbf{n})$  of complex numbers indexed by  $\mathbf{n} \in \mathbb{N}^k$ , we will define partial summation of  $a(\mathbf{n})$  with respect to such partitions as follows:

$$A_{(I),(J)}(\mathbf{t}_{(I)}, \mathbf{s}_{(J)}) := \sum_{\substack{0 < n_j \leq t_j \\ j \in I}} \sum_{\substack{0 < n_j \leq s_j \\ j \in J}} a(\mathbf{n}).$$

More specifically, in our application, given a partition  $I \cup J$  of  $\{1, \dots, k\}$ , a tuple  $\mathbf{m}$  of integers and tuples  $\mathbf{s}, \mathbf{t} \in \mathbb{R}^k$  of positive real numbers and a fixed index  $\alpha$  of a box  $B_\alpha$ , we define

$$T_{(I),(J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{t}_{(I)}, \mathbf{s}_{(J)}) := \left| \sum_{\substack{0 < n_j \leq t_j \\ j \in I}} \sum_{\substack{0 < n_j \leq s_j \\ j \in J}} e(\theta_\alpha(\mathbf{n})) \left( \prod_{i=1}^k \chi_i(n_i + m_i) \right) \right|.$$

The key approximation lemma is as follows:

**Lemma 4.1.** *Given integral tuples  $\mathbf{m}$  and  $\mathbf{Q}$  with*

$$(4.5) \quad Q_i \geq K_i \quad \text{for each } i = 1, \dots, k,$$

*the above dissection provides an index  $\alpha$  such that*

$$(4.6) \quad T_0(\mathbf{F}; \mathbf{m}, \mathbf{K}) \ll_{k,d} \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} K_j^{-1} \right) \int \cdots \int_{(0, K_j]} T_{(cJ), (J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K}_{(cJ)}, \mathbf{t}_{(J)}) d\mathbf{t}_{(J)}.$$

*Here the sum is over all subsets  $J \subseteq \{1, \dots, k\}$ , with corresponding complement  ${}^cJ = \{1, \dots, k\} \setminus J$ . If  $J = \{j_1, \dots, j_v\}$  then we set  $d\mathbf{t}_{(J)} = dt_{j_1} \cdots dt_{j_v}$ .*

To prove this, we first observe that for an integral tuple  $\mathbf{m}$ ,

$$\begin{aligned} T_0(\mathbf{F}; \mathbf{m}, \mathbf{K}) &= \sup_{f \in \mathcal{F}_0(\mathbf{F})} \left| \sum_{\mathbf{m} < \mathbf{x} \leq \mathbf{m} + \mathbf{K}} e(f(\mathbf{x})) \chi_1(x_1) \cdots \chi_k(x_k) \right| \\ &= \sup_{f \in \mathcal{F}_0(\mathbf{F})} \left| \sum_{\mathbf{0} < \mathbf{x} \leq \mathbf{K}} e(f(\mathbf{x})) \chi_1(x_1 + m_1) \cdots \chi_k(x_k + m_k) \right|. \end{aligned}$$

We now write

$$T_0(\mathbf{F}; \mathbf{m}, \mathbf{K}) = \left| \sum_{\mathbf{0} < \mathbf{x} \leq \mathbf{K}} e(f(\mathbf{x})) \chi_1(x_1 + m_1) \cdots \chi_k(x_k + m_k) \right|,$$

for some fixed polynomial  $f \in \mathcal{F}_0(\mathbf{F})$ , which we write explicitly as

$$f(\mathbf{X}) = \sum_{\beta \in \Lambda_0(\mathbf{F})} f_\beta \mathbf{X}^\beta,$$

where as before we may assume that each  $f_\beta \in [0, 1]$ . Given our dissection of the coefficient space  $[0, 1]^{R+1}$ , we may choose a box  $B_\alpha$  with index  $\alpha$  and distinguished vertex  $\theta_\alpha = (\theta_{\alpha, \beta})_\beta$

such that

$$(4.7) \quad |f_\beta - \theta_{\alpha,\beta}| \leq \mathbf{Q}^{-\beta}, \quad \text{for each multi-index } \beta \in \Lambda_0(\mathbf{F}).$$

(This is simply choosing  $\alpha$  such that the coefficient tuple  $(f_\beta)_\beta$  lies in the box  $B_\alpha$ .) For notational convenience, we will temporarily set  $\delta_\beta = f_\beta - \theta_{\alpha,\beta}$  for each  $\beta \in \Lambda_0(\mathbf{F})$ . We then write

$$(4.8) \quad \sum_{\mathbf{0} < \mathbf{x} \leq \mathbf{K}} e(f(\mathbf{x})) \chi_1(x_1 + m_1) \cdots \chi_k(x_k + m_k) \\ = \sum_{\mathbf{0} < \mathbf{x} \leq \mathbf{K}} e \left( \sum_{\beta \in \Lambda_0(\mathbf{F})} \delta_\beta \mathbf{x}^\beta \right) e(\theta_\alpha(\mathbf{x})) \chi_1(x_1 + m_1) \cdots \chi_k(x_k + m_k).$$

We now apply summation by parts, in the following form, which we prove in Section 7.3:

**Lemma 4.2.** *Let  $a(\mathbf{n})$  be a sequence of complex numbers indexed by  $\mathbf{n} \in \mathbb{Z}^k \cap (\mathbf{0}, \mathbf{N}]$ . Let  $b(\mathbf{x})$  be a  $C^{(k)}$  function on  $\mathbb{R}^k$  such that there is a tuple  $\mathbf{B} = (B_1, \dots, B_k)$  of positive real numbers such that for every multi-index*

$$(4.9) \quad \kappa = (\kappa_1, \dots, \kappa_k) \quad \text{with } \kappa_i \in \{0, 1\}$$

we have

$$(4.10) \quad \left| \frac{\partial^{|\kappa|}}{\partial x_1^{\kappa_1} \cdots \partial x_k^{\kappa_k}} b(\mathbf{x}) \right| \leq B_1^{\kappa_1} \cdots B_k^{\kappa_k} = \mathbf{B}^\kappa \quad \text{for all } \mathbf{x} \in (\mathbf{0}, \mathbf{N}].$$

Then

$$\left| \sum_{\mathbf{n} \leq \mathbf{N}} a(\mathbf{n}) b(\mathbf{n}) \right| \leq \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} B_j \right) \int \cdots \int_{(0, N_j]} |A_{({}^c J), (J)}(\mathbf{N}_{({}^c J)}, \mathbf{t}_{(J)})| d\mathbf{t}_{(J)}.$$

Here the sum is over all subsets  $J \subseteq \{1, \dots, k\}$ , with corresponding complement  ${}^c J = \{1, \dots, k\} \setminus J$ . If  $J = \{j_1, \dots, j_v\}$  then we set  $d\mathbf{t}_{(J)} = dt_{j_1} \cdots dt_{j_v}$ .

(Note that if  $\kappa = (0, 0, \dots, 0)$  then (4.10) is simply the assumption that  $|b(\mathbf{x})| \leq 1$ .) We apply this lemma to (4.8) with the choices  $\mathbf{N} = \mathbf{K}$  and

$$a(\mathbf{x}) = e(\theta_\alpha(\mathbf{x})) \chi_1(x_1 + m_1) \cdots \chi_k(x_k + m_k), \\ b(\mathbf{x}) = e \left( \sum_{\beta \in \Lambda_0(\mathbf{F})} \delta_\beta \mathbf{x}^\beta \right).$$



We may verify that for a fixed index  $j$ , if we let  $e_j = (0, \dots, 1, \dots, 0)$  be the  $j$ -th unit multi-index, then for  $\mathbf{x} \in (\mathbf{0}, \mathbf{K}]$ ,

$$\begin{aligned}
\left| \frac{\partial}{\partial x_j} e \left( \sum_{\beta \in \Lambda_0(\mathbf{F})} \delta_\beta \mathbf{x}^\beta \right) \right| &= \left| \frac{\partial}{\partial x_j} \left( 2\pi i \sum_{\beta \in \Lambda_0(\mathbf{F})} \delta_\beta \mathbf{x}^\beta \right) \right| \\
&\leq 2\pi \sum_{\substack{\beta \in \Lambda_0(\mathbf{F}) \\ \beta_j \geq 1}} \beta_j |\delta_\beta| |\mathbf{x}^{\beta - e_j}| \\
&\leq 2\pi \sum_{\substack{\beta \in \Lambda_0(\mathbf{F}) \\ \beta_j \geq 1}} \beta_j |\delta_\beta| |\mathbf{K}^{\beta - e_j}| \\
&\leq 2\pi \sum_{\substack{\beta \in \Lambda_0(\mathbf{F}) \\ \beta_j \geq 1}} \beta_j \mathbf{Q}^{-\beta} \mathbf{K}^{\beta - e_j} \\
&\ll_{k,d} K_j^{-1},
\end{aligned}$$

where we have used the assumption (4.7) on the size of  $\delta_\beta$ , followed by the assumption (4.5) that  $Q_i \geq K_i$ .

Similarly, one may compute that for each fixed  $\kappa$  of the form (4.9),

$$\left| \frac{\partial^{|\kappa|}}{\partial x_1^{\kappa_1} \dots \partial x_k^{\kappa_k}} e \left( \sum_{\beta \in \Lambda_0(\mathbf{F})} \delta_\beta \mathbf{x}^\beta \right) \right| \ll_{k,d} \mathbf{K}^{-\kappa},$$

so that a bound of the form (4.10) is satisfied with  $B_i = K_i^{-1}$ . We may thus apply Lemma 4.2 to (4.8) to conclude that (4.6) holds.

A repeated application of Hölder's inequality in (4.6) shows that  $T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r}$  is at most

$$\ll_{k,r,d} \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} K_j^{-1} \right) \int \dots \int_{\substack{(0, K_j] \\ j \in J}} T_{(cJ), (J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K}_{(cJ)}, \mathbf{t}_{(J)})^{2r} d\mathbf{t}_{(J)}.$$

This is still for the fixed index  $\alpha$  provided by Lemma 4.1; as in [13], in order to eliminate the awkward dependence on  $\alpha$ , we sum trivially on the right hand side over all values of the parameter  $\alpha$  that indexes the boxes in the dissection (4.1), so that by positivity,  $T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r}$  is at most

$$\ll_{k,r,d} \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} K_j^{-1} \right) \int \dots \int_{\substack{(0, K_j] \\ j \in J}} \sum_{\alpha} T_{(cJ), (J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K}_{(cJ)}, \mathbf{t}_{(J)})^{2r} d\mathbf{t}_{(J)}.$$

Now we sum over  $\mathbf{m} \pmod{\mathbf{q}}$ , so that  $\sum_{\mathbf{m} \pmod{\mathbf{q}}} T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r}$  is at most

$$\ll \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} K_j^{-1} \right) \int \dots \int_{\substack{(0, K_j] \\ j \in J}} \sum_{\alpha} \sum_{\mathbf{m} \pmod{\mathbf{q}}} T_{(cJ), (J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K}_{(cJ)}, \mathbf{t}_{(J)})^{2r} d\mathbf{t}_{(J)}.$$

We note that if  $\mathbf{t} \leq \mathbf{K}$  then by positivity, for any index set  $J$ ,

$$(4.11) \quad \sum_{\alpha} \sum_{\mathbf{m} \pmod{\mathbf{q}}} T_{(cJ), (J)}(\alpha, \mathbf{F}; \mathbf{m}, \mathbf{K}_{(cJ)}, \mathbf{t}_{(J)})^{2r} \leq \sup_{\tau \leq \mathbf{K}} \left( \sum_{\alpha} \sum_{\mathbf{m} \pmod{\mathbf{q}}} T(\alpha, \mathbf{F}; \mathbf{m}, \tau)^{2r} \right),$$

where  $T(\alpha, \mathbf{F}; \mathbf{m}, \tau)$  is defined by (4.4). Applying this in the integrand above and noting the normalization complementing the region of integration, we see that

$$\sum_{\mathbf{m} \pmod{\mathbf{q}}} T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r} \ll_{k,r,d} \sup_{\tau \leq \mathbf{K}} \left( \sum_{\alpha} \sum_{\mathbf{m} \pmod{\mathbf{q}}} T(\alpha, \mathbf{F}; \mathbf{m}, \tau)^{2r} \right).$$

For convenience, we define

$$S_4(\tau) = \sum_{\alpha} \sum_{\mathbf{m} \pmod{\mathbf{q}}} T(\alpha, \mathbf{F}; \mathbf{m}, \tau)^{2r}.$$

We may conclude:

**Lemma 4.3.**

$$(4.12) \quad S_3(\mathbf{K}) = \sum_{\mathbf{m} \pmod{\mathbf{q}}} T_0(\mathbf{F}; \mathbf{m}, \mathbf{K})^{2r} \ll_{k,r,d} \sup_{\tau \leq \mathbf{K}} S_4(\tau).$$

We now summarize what we have proved so far, by combining the result of Lemma 4.3 with (3.5) and (3.6):

**Proposition 4.4.** *As long as  $\mathbf{Q} \geq \mathbf{K}$ ,  $\mathbf{K} = 2\mathbf{H}/\mathbf{P} < \mathbf{q}$ , and  $H_i P_i < q_i$  for all  $i = 1, \dots, k$ ,*

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \mathcal{L}(\mathbf{q})^2 \left\{ \sup_{\tau \leq 2\mathbf{H}/\mathbf{P}} S_4(\tau) \right\}^{\frac{1}{2r}}.$$

Thus our focus turns to bounding  $S_4(\tau)$  for any fixed tuple  $\tau$  with  $\tau \leq \mathbf{K} = 2\mathbf{H}/\mathbf{P}$ . We recall the definition of the boxes  $B_\alpha$ , the vertex  $\theta_\alpha$  associated to each box  $B_\alpha$ , and the associated polynomial  $\theta_\alpha(\mathbf{x})$  defined by (4.3). We will represent a set of cardinality  $2r$  of tuples  $\mathbf{x}^{(j)} = (x_1^{(j)}, \dots, x_k^{(j)}) \in \mathbb{Z}^k$  by  $\{\mathbf{x}\} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)}\}$ . We set

$$\Sigma_A(\{\mathbf{x}\}) := \sum_{\alpha} e \left( \sum_{j=1}^{2r} \varepsilon(j) \theta_\alpha(\mathbf{x}^{(j)}) \right),$$

where  $\varepsilon(j) := (-1)^j$ . We note that we may trivially bound  $\Sigma_A$  by the number of summands, that is the number of boxes, namely

$$(4.13) \quad |\Sigma_A(\{\mathbf{x}\})| \leq \mathbf{Q}^\gamma,$$

where  $\gamma = \gamma(\mathbf{F})$  is the density of the system  $\mathbf{F}$ .

For each  $i = 1, \dots, k$  let  $\Delta_i(q_i)$  denote the order of  $\chi_i$  modulo  $q_i$ ; furthermore for each  $j = 1, \dots, 2r$  set  $\delta_i(j) = 1$  if  $j$  is even and  $\delta_i(j) = \Delta_i(q_i) - 1$  if  $j$  is odd. Now define for each  $i = 1, \dots, k$  the single-variable polynomial

$$G_i(\Delta_i(q_i), \{\mathbf{x}\}; X) := \prod_{j=1}^{2r} (X + x_i^{(j)})^{\delta_i(j)}.$$

Finally, set

$$\Sigma_B(\{\mathbf{x}\}; \mathbf{q}) := \prod_{i=1}^k \left( \sum_{m_i=1}^{q_i} \chi_i(G_i(\Delta_i(q_i), \{\mathbf{x}\}; m_i)) \right).$$

We now expand the sums in the definition of  $S_4(\boldsymbol{\tau})$  to see that with this notation,

$$(4.14) \quad S_4(\boldsymbol{\tau}) = \sum_{\substack{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \\ \mathbf{0} < \mathbf{x}^{(j)} \leq \boldsymbol{\tau}}} \Sigma_A(\{\mathbf{x}\}) \Sigma_B(\{\mathbf{x}\}; \mathbf{q}).$$

We will now proceed in two parts: first, we will sum trivially over the boxes  $B_\alpha$ , using the trivial bound (4.13) for  $\Sigma_A$ . This will result in the following proposition:

**Proposition 4.5.** *Under the assumption that the tuple  $\mathbf{K}$  satisfies*

$$q_i^{\frac{1}{2r}} \ll K_i \ll q_i^{\frac{1}{2r}} \quad \text{for each } i = 1, \dots, k,$$

*we have*

$$\sup_{\boldsymbol{\tau} \leq \mathbf{K}} S_4(\boldsymbol{\tau}) \ll \mathbf{Q}^\gamma \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{\frac{1}{2}},$$

*where  $\gamma$  is the density of the system  $\mathbf{F}$  and the implied constant depends on  $r, d, k$ .*

As an immediate consequence we will prove:

**Theorem 4.6.** *Let  $\mathbf{F}$  be a reduced monomial translation-dilation invariant system having dimension  $k$ , degree  $d$ , rank  $R$ , weight  $M$  and density  $\gamma$ . Let  $\mathcal{F}$  denote the set of all real-valued polynomials spanned by the system  $\mathbf{F}$ . If  $r \geq 1$  and  $q_i^{\frac{1}{2r}} < H_i < q_i^{\frac{1}{2} + \frac{1}{4r}}$  for each  $i = 1, \dots, k$ , then*

$$\sup_{f \in \mathcal{F}} |S_k(f; \mathbf{N}, \mathbf{H})| \ll_{r,d,k} \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{r+1}{4r^2}} (\mathbf{q}^\gamma)^{\frac{1}{4r^2}} \mathcal{L}(\mathbf{q})^2.$$

This is the result that leads to (2.6) in the case  $q_1 = \dots = q_k = q$ . We will improve on this in Section 6 by proving a nontrivial upper bound for  $\Sigma_A$  via Theorem B, which we will consequently use to give a refinement (Proposition 6.2) of Proposition 4.5. Our main general result, Theorem 2.1, will then follow from this refinement. For purposes of comparison, we note that when for example  $q_1 = \dots = q_k = q$ , Theorem 2.1 is sharper than Theorem 4.6 for  $r > M + k$ , as well as in the sense that (2.5) improves on (2.6).

## 5. THE MULTIPLICATIVE COMPONENT

Our treatment of the multiplicative component  $\Sigma_B$  is the same for both Theorem 4.6 and Theorem 2.1, and hinges upon an application of the Weil bound. It will be convenient to regard a collection  $\{\mathbf{x}\}$  as either a set of cardinality  $2r$  of  $k$ -tuples  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k$ , or equivalently as a set of cardinality  $k$  of  $2r$ -tuples, which we will denote by  $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)} \in \mathbb{Z}^{2r}$ ; in matrix form we are regarding the  $\mathbf{x}^{(j)}$  as the rows of a  $2r \times k$  matrix, and the  $\mathbf{z}^{(i)}$  as the columns. We will denote such a collection  $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)}$  also by  $\{\mathbf{z}\}$ .

We recall the definitions of  $\Delta_i(q_i)$  and  $\delta_i(j)$ . We now define for any  $2r$ -tuple  $\mathbf{z} = (z_1, \dots, z_{2r})$  the single-variable polynomial

$$F(\Delta_i(q_i), \mathbf{z}; X) := \prod_{j=1}^{2r} (X + z_j)^{\delta_i(j)}.$$

We can now write

$$(5.1) \quad \Sigma_B(\{\mathbf{x}\}; \mathbf{q}) = \Sigma_B(\{\mathbf{z}\}; \mathbf{q}) = \prod_{i=1}^k \Sigma_B(\mathbf{z}^{(i)}; \chi_i, q_i)$$

where we define

$$\Sigma_B(\mathbf{z}^{(i)}; \chi_i, q_i) := \sum_{m_i=1}^{q_i} \chi_i \left( F(\Delta_i(q_i), \mathbf{z}^{(i)}; m_i) \right).$$

We aim to apply the following consequence of the well-known Weil bound:

**Lemma 5.1.** *Let  $\chi$  be a character of order  $\Delta(q) > 1$  modulo a prime  $q$ . Suppose that  $F(X)$  is a polynomial which is not a perfect  $\Delta(q)$ -th power in  $\mathbb{F}_q[X]$ . Then*

$$\left| \sum_{m=1}^q \chi(F(m)) \right| \leq (\deg(F) - 1) \sqrt{q}.$$

For a fixed  $i$ , we can apply Lemma 5.1 to bound

$$(5.2) \quad \Sigma_B(\mathbf{z}^{(i)}; \chi_i, q_i) \ll q_i^{1/2},$$

unless  $F(\Delta_i(q_i), \mathbf{z}^{(i)}; X)$  is a perfect  $\Delta_i(q_i)$ -th power over  $\mathbb{F}_{q_i}$ , in which case we apply the trivial bound

$$(5.3) \quad \Sigma_B(\mathbf{z}^{(i)}; \chi_i, q_i) \ll q_i.$$

We define a  $2r$ -tuple  $\mathbf{z} = (z_1, \dots, z_{2r})$  to be *bad* if for each  $j = 1, \dots, 2r$  there exists  $\ell \neq j$  such that  $z_\ell = z_j$ , and to be *good* otherwise. We have the following simple statement:

**Lemma 5.2.** *Fix a character  $\chi$  with order  $\Delta(q) > 1$  modulo a prime  $q$ . Fix a tuple  $\mathbf{z} = (z_1, \dots, z_{2r})$  with  $0 < z_j \leq u$  for each  $j = 1, \dots, 2r$ . If  $u \leq q$  and  $F(\Delta(q), \mathbf{z}; X)$  is a perfect  $\Delta(q)$ -th power modulo  $q$ , then  $\mathbf{z}$  is bad.*

This result is clear, since if  $\mathbf{z}$  were good, there would be a value  $y$ , say, which is taken only by  $z_j$  for one index  $j \in \{1, \dots, 2r\}$ ; thus the factor  $(X + y)$  would appear with multiplicity 1 or  $\Delta(q) - 1$  in  $F(\Delta(q), \mathbf{z}; X)$ , neither of which is divisible by  $\Delta(q)$ .

Lemma 5.2 is useful for a single factor  $\Sigma_B(\mathbf{z}^{(i)}; \chi_i, q_i)$ , but we must also consider how many tuples in a collection  $\{\mathbf{z}\}$  are bad. For each subset  $\mathcal{S} \subseteq \{1, \dots, k\}$  (possibly empty), we say a collection  $\{\mathbf{z}\} = \{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)}\}$  is  $\mathcal{S}$ -bad if  $\mathbf{z}^{(i)}$  is bad for  $i \in \mathcal{S}$  and good for  $i \notin \mathcal{S}$ . For each subset  $\mathcal{S} \subseteq \{1, \dots, k\}$ , we let  $\mathcal{B}(\mathcal{S}; \boldsymbol{\tau})$  denote the set of collections  $\{\mathbf{z}\}$  that are  $\mathcal{S}$ -bad and such that for each  $1 \leq i \leq k$  the entries of  $\mathbf{z}^{(i)}$  are at most  $\tau_i$ . (Implicitly this also specifies when the original tuple  $\{\mathbf{x}\}$  belongs to  $\mathcal{B}(\mathcal{S}; \boldsymbol{\tau})$ ; we will use this in the computation (6.4).) We now prove an upper bound on the cardinality of the set  $\mathcal{B}(\mathcal{S}; \boldsymbol{\tau})$ :

**Lemma 5.3.** *For any fixed subset  $\mathcal{S} \subseteq \{1, \dots, k\}$ ,*

$$\#\mathcal{B}(\mathcal{S}; \boldsymbol{\tau}) \leq r^{k(2r+1)} \left( \prod_{i \in \mathcal{S}} \tau_i^r \right) \left( \prod_{i \notin \mathcal{S}} \tau_i^{2r} \right).$$

We recall from the classical Burgess method (see for example Lemma 3.2 of [13]) that there are at most  $r^{2r+1}u^r$  choices for a single bad  $2r$ -tuple  $\mathbf{z}$  with entries at most  $u$ . Fix a subset  $\mathcal{S} \subseteq \{1, \dots, k\}$ . For each  $i \in \mathcal{S}$  there are then at most  $r^{2r+1}\tau_i^r$  choices for  $\mathbf{z}^{(i)}$  that are bad, and for each  $i \notin \mathcal{S}$  there are trivially at most  $\tau_i^{2r}$  ways to choose  $\mathbf{z}^{(i)}$  that are good. Thus upon recalling  $|\mathcal{S}| \leq k$ , the lemma is proved.

Finally, we note:

**Lemma 5.4.** *Suppose  $\{\mathbf{z}\} \in \mathcal{B}(\mathcal{S}, \boldsymbol{\tau})$  and let  $\sigma(\mathcal{S})$  be the indicator multi-index for  $\mathcal{S}$ , that is  $\sigma(\mathcal{S}) = (\sigma_1, \dots, \sigma_k)$  with  $\sigma_i = 1$  if  $i \in \mathcal{S}$  and  $\sigma_i = 0$  if  $i \notin \mathcal{S}$ . If  $\boldsymbol{\tau}$  is such that  $\tau_i \leq q_i$  for each  $i = 1, \dots, k$ , then*

$$(5.4) \quad \Sigma_B(\{\mathbf{z}\}; \mathbf{q}) \ll_{r,k} \|\mathbf{q}\|^{\frac{1}{2}} \mathbf{q}^{\sigma(\mathcal{S})/2}.$$

We simply note that within the product (5.1) we may apply the Weil bound (5.2) for each  $i \notin \mathcal{S}$  and the trivial bound (5.3) for each  $i \in \mathcal{S}$ ; this suffices for the lemma.

We now consider (4.14), applying the trivial bound (4.13) to  $\Sigma_A$  and decomposing  $\Sigma_B$  as follows:

$$\begin{aligned} S_4(\boldsymbol{\tau}) &= \sum_{\substack{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \\ \mathbf{0} < \mathbf{x}^{(j)} \leq \boldsymbol{\tau}}} \Sigma_A(\{\mathbf{x}\}) \Sigma_B(\{\mathbf{x}\}; \mathbf{q}) \\ &\leq \mathbf{Q}^\gamma \sum_{\substack{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \\ \mathbf{0} < \mathbf{x}^{(j)} \leq \boldsymbol{\tau}}} |\Sigma_B(\{\mathbf{x}\}; \mathbf{q})| \\ &= \mathbf{Q}^\gamma \sum_{\mathcal{S} \subseteq \{1, \dots, k\}} \sum_{\{\mathbf{z}\} \in \mathcal{B}(\mathcal{S}; \boldsymbol{\tau})} |\Sigma_B(\{\mathbf{z}\}; \mathbf{q})| \\ &\ll_{r,k} \mathbf{Q}^\gamma \sum_{\mathcal{S} \subseteq \{1, \dots, k\}} \left( \prod_{i \in \mathcal{S}} \tau_i^r \right) \left( \prod_{i \notin \mathcal{S}} \tau_i^{2r} \right) \|\mathbf{q}\|^{1/2} \mathbf{q}^{\sigma(\mathcal{S})/2}. \end{aligned}$$

Here we have applied Lemmas 5.3 and 5.4. We now note that since this is monotone in each  $\tau_i$ , we have the upper bound

$$\sup_{\boldsymbol{\tau} \leq \mathbf{K}} S_4(\boldsymbol{\tau}) \ll \mathbf{Q}^\gamma \sum_{\mathcal{S} \subseteq \{1, \dots, k\}} \left( \prod_{i \in \mathcal{S}} K_i^r \right) \left( \prod_{i \notin \mathcal{S}} K_i^{2r} \right) \|\mathbf{q}\|^{1/2} \mathbf{q}^{\sigma(\mathcal{S})/2}.$$

We re-write this as

$$\sup_{\boldsymbol{\tau} \leq \mathbf{K}} S_4(\boldsymbol{\tau}) \ll \mathbf{Q}^\gamma \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{1/2} \left\{ 1 + \sum_{\substack{\mathcal{S} \subseteq \{1, \dots, k\} \\ \mathcal{S} \neq \emptyset}} \mathbf{K}^{-r\sigma(\mathcal{S})} \mathbf{q}^{\sigma(\mathcal{S})/2} \right\}.$$

Under the assumption

$$(5.5) \quad q_i^{\frac{1}{2r}} \ll K_i \ll q_i^{\frac{1}{2r}} \quad \text{for each } i = 1, \dots, k,$$

we have  $\mathbf{K}^{-r\sigma(\mathcal{S})} \mathbf{q}^{\sigma(\mathcal{S})/2} = O(1)$  for each subset  $\mathcal{S}$  and as a result

$$\sup_{\boldsymbol{\tau} \leq \mathbf{K}} S_4(\boldsymbol{\tau}) \ll_{r,k} \mathbf{Q}^\gamma \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{\frac{1}{2}},$$

which proves Proposition 4.5.

**5.1. Proof of Theorem 4.6.** With Proposition 4.5 in hand, it is simple to derive Theorem 4.6. Recalling that  $\mathbf{K} = 2\mathbf{H}/\mathbf{P}$ , the condition (5.5) leads us to choose the parameters  $P_i$  such that

$$\frac{1}{2} H_i q_i^{-\frac{1}{2r}} \leq P_i < H_i q_i^{-\frac{1}{2r}}, \quad \text{for each } i = 1, \dots, k,$$

in which case (5.5) holds. We also note that as long as  $H_i > q_i^{\frac{1}{2r}}$ , we may choose  $P_i \geq 1$ . We furthermore choose  $Q_i = \lceil K_i \rceil$  for each  $i$ . In order to satisfy the further conditions  $H_i P_i < q_i$  of Lemma 3.2, we must restrict to ranges with  $H_i < q_i^{\frac{1}{2} + \frac{1}{4r}}$ .

With these choices, we input the result of Proposition 4.5 into Proposition 4.4 to see that

$$\begin{aligned} T(\mathbf{F}; \mathbf{N}, \mathbf{H}) &\ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \mathcal{L}(\mathbf{q})^2 \left\{ \mathbf{Q}^\gamma \|\mathbf{H}\|^{2r} \|\mathbf{P}\|^{-2r} \|\mathbf{q}\|^{\frac{1}{2}} \right\}^{\frac{1}{2r}} \\ &\ll \|\mathbf{H}\|^{1-\frac{1}{2r}} \|\mathbf{P}\|^{-\frac{1}{2r}} (\mathbf{H}/\mathbf{P})^{\frac{\gamma}{2r}} \|\mathbf{q}\|^{\frac{1}{4r}} \mathcal{L}(\mathbf{q})^2. \end{aligned}$$

We now note that because of our choice of  $\mathbf{P}$ ,

$$(\mathbf{H}/\mathbf{P})^{\frac{\gamma}{2r}} \ll (\mathbf{q}^\gamma)^{\frac{1}{4r^2}}.$$

Thus we may conclude that

$$\begin{aligned} T(\mathbf{F}; \mathbf{N}, \mathbf{H}) &\ll \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{1}{4r^2}} (\mathbf{q}^\gamma)^{\frac{1}{4r^2}} \|\mathbf{q}\|^{\frac{1}{4r}} \mathcal{L}(\mathbf{q})^2 \\ &\ll \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{r+1}{4r^2}} (\mathbf{q}^\gamma)^{\frac{1}{4r^2}} \mathcal{L}(\mathbf{q})^2, \end{aligned}$$

which proves Theorem 4.6. In particular, if  $q_j = q$  for all  $j$ , this simplifies to

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1-\frac{1}{r}} q^{\frac{k(r+1)+M}{4r^2}} (\log q)^{2k},$$

where we recall that  $M = M(\mathbf{F})$  is the weight of the system  $\mathbf{F}$ .

**5.2. Optimal choice of  $r$ .** We make a remark on the case  $q_1 = \dots = q_k = q$  and the optimal choice of  $r$ . Suppose that for each  $i$ ,  $H_i = q^{1/4+\kappa_i}$ . Set  $\sigma = \sum_{i=1}^k \kappa_i$ . Then Theorem 4.6 provides an upper bound of the size  $\|\mathbf{H}\| q^{-\delta+\varepsilon}$  where  $\delta = (\sigma r - \frac{1}{4}(k+M))r^{-2}$ . As a function of  $r$ , this attains a maximum at the real value  $r_0 = (k+M)(2\sigma)^{-1}$ . Choosing  $r = r_0 + \theta$  where  $-1/2 \leq \theta < 1/2$ , we see that as claimed in (2.6),  $\delta$  is approximately of size  $\delta \approx \sigma^2(M+k)^{-1}$ .

## 6. THE ADDITIVE COMPONENT: NONTRIVIAL ANALYSIS

We now return to a nontrivial analysis of the additive component  $\Sigma_A$ , which will lead to our main result Theorem 2.1. Our goal is to connect the analysis of  $\Sigma_A$  to a Vinogradov Mean Value Theorem for the translation-dilation invariant system  $\mathbf{F}$ . We again recall the definition of the boxes  $B_\alpha$  that partition the coefficient space  $[0, 1]^{R+1}$ , and in particular the definition (4.2) of the distinguished vertex  $\theta_\alpha$  associated to each box  $B_\alpha$ . It is convenient

to recall the fixed ordering  $\beta^{(0)}, \dots, \beta^{(R)}$  of the multi-indices  $\beta \in \Lambda_0(\mathbf{F})$ . We compute:

$$\begin{aligned}
\Sigma_A(\{\mathbf{x}\}) &= \sum_{\alpha} e \left( \sum_{j=1}^{2r} \varepsilon(j) \theta_{\alpha}(\mathbf{x}^{(j)}) \right) \\
&= \sum_{\alpha} e \left( \sum_{\beta \in \Lambda_0(\mathbf{F})} \theta_{\alpha, \beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \\
&= \sum_{\alpha} e \left( \sum_{\beta = \beta^{(0)}, \dots, \beta^{(R)}} \theta_{\alpha, \beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \\
&= \sum_{c_{\beta^{(0)}}, \dots, c_{\beta^{(R)}}} e \left( \sum_{\beta = \beta^{(0)}, \dots, \beta^{(R)}} c_{\beta} \mathbf{Q}^{-\beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right),
\end{aligned}$$

where the sum over  $c_{\beta^{(0)}}, \dots, c_{\beta^{(R)}}$  indicates summing for each  $i = 0, \dots, R$  the parameter  $c_{\beta^{(i)}}$  over integers  $0 \leq c_{\beta^{(i)}} \leq \mathbf{Q}^{\beta^{(i)}} - 1$ . Thus

$$\begin{aligned}
\Sigma_A(\{\mathbf{x}\}) &= \sum_{c_{\beta^{(0)}}, \dots, c_{\beta^{(R)}}} \left\{ \prod_{\beta = \beta^{(0)}, \dots, \beta^{(R)}} e \left( c_{\beta} \mathbf{Q}^{-\beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \right\} \\
&= \prod_{\beta = \beta^{(0)}, \dots, \beta^{(R)}} \left\{ \sum_{c_{\beta} \pmod{\mathbf{Q}^{\beta}}} e \left( c_{\beta} \mathbf{Q}^{-\beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \right\} \\
&= \prod_{\beta \in \Lambda_0(\mathbf{F})} \left\{ \sum_{c_{\beta} \pmod{\mathbf{Q}^{\beta}}} e \left( c_{\beta} \mathbf{Q}^{-\beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \right\}.
\end{aligned}$$

Since the multi-index  $\beta = (0, \dots, 0)$  contributes only a multiplicative factor of 1, we see that

$$\Sigma_A(\{\mathbf{x}\}) = \prod_{\beta \in \Lambda(\mathbf{F})} \left\{ \sum_{c_{\beta} \pmod{\mathbf{Q}^{\beta}}} e \left( c_{\beta} \mathbf{Q}^{-\beta} \left( \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \right) \right) \right\}.$$

By orthogonality of characters we therefore have

$$\Sigma_A(\{\mathbf{x}\}) = \mathbf{Q}^{\gamma} \Xi_{\mathbf{Q}}(\mathbf{F}; \{\mathbf{x}\})$$

where  $\Xi_{\mathbf{Q}}(\mathbf{F}; \{\mathbf{x}\})$  is the indicator function for the set

$$(6.1) \quad \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \cap (0, \tau] : \sum_{j=1}^{2r} \varepsilon(j) (\mathbf{x}^{(j)})^{\beta} \equiv 0 \pmod{\mathbf{Q}^{\beta}}, \forall \beta \in \Lambda(\mathbf{F})\}.$$

Here we have also used the fact, previously observed, that

$$\prod_{\beta \in \Lambda(\mathbf{F})} \mathbf{Q}^{\beta} = \mathbf{Q}^{\gamma}.$$

In our application we will have  $\tau_i \leq K_i = 2H_i/P_i$  for each  $i = 1, \dots, k$ . So far we have only assumed that  $Q_i \geq K_i$  for each  $i$ ; we now furthermore assume that each  $Q_i$  is sufficiently

large that the congruences in the definition of the set (6.1) must be identities in  $\mathbb{Z}$ . We check that for any multi-index  $\beta \in \Lambda(\mathbf{F})$  and any collection  $\{\mathbf{x}\}$  in the set (6.1),

$$\left| \sum_{j=1}^{2r} \varepsilon(j)(\mathbf{x}^{(j)})^\beta \right| < 2r\tau^\beta \leq 2r\mathbf{K}^\beta \leq (2r\mathbf{K})^\beta = (4r\mathbf{H}/\mathbf{P})^\beta.$$

Thus we choose

$$(6.2) \quad Q_i = \lceil 4rH_i/P_i \rceil \quad \text{for each } i = 1, \dots, k.$$

With this choice the congruences in (6.1) must be identities in  $\mathbb{Z}$ , and we may replace  $\Xi_{\mathbf{Q}}(\mathbf{F}; \{\mathbf{x}\})$  by the indicator function  $\Xi(\mathbf{F}; \{\mathbf{x}\})$  of the set

$$V_r(\mathbf{F}; \tau) := \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \cap (\mathbf{0}, \tau] : \sum_{j=1}^{2r} \varepsilon(j)(\mathbf{x}^{(j)})^\beta = 0, \forall \beta \in \Lambda(\mathbf{F})\}.$$

We have shown:

**Proposition 6.1.** *Given a collection  $\{\mathbf{x}\} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)}\}$  with  $\mathbf{x}^{(j)} \in \mathbb{Z}^k \cap (\mathbf{0}, \tau]$  for each  $j = 1, \dots, 2r$ , if  $\tau \leq \mathbf{K} = 2\mathbf{H}/\mathbf{P}$  and we choose  $\mathbf{Q}$  as in (6.2), then*

$$\Sigma_A(\{\mathbf{x}\}) = \mathbf{Q}^\gamma \Xi(\mathbf{F}; \{\mathbf{x}\}).$$

We now set

$$\tau_{\max} = \max\{\tau_1, \dots, \tau_k\},$$

and note that

$$\#V_r(\mathbf{F}; \tau) \leq J_r(\mathbf{F}; \tau_{\max}),$$

where  $J_r(\mathbf{F}; X)$  is the counting function for the system of equations (2.1) corresponding to the given reduced monomial translation-dilation invariant system  $\mathbf{F}$ . We recall from (2.4) that the conjectured upper bound is

$$(6.3) \quad J_r(\mathbf{F}; X) \ll X^{2rk-M+\varepsilon},$$

known unconditionally for  $r \geq R(d+1)$  by Theorem B (Section 2.1).

We now return to the consideration of  $S_4(\tau)$ , given in (4.14) in terms of the additive component  $\Sigma_A$  and the multiplicative component  $\Sigma_B$ . Define for each subset  $\mathcal{S} \subseteq \{1, \dots, k\}$  the quantity

$$N(\mathcal{S}; \tau) = \#\{\mathcal{B}(\mathcal{S}; \tau) \cap V_r(\mathbf{F}; \tau)\}.$$

With  $\tau \leq \mathbf{K}$  and  $\mathbf{Q}$  as above, we apply Proposition 6.1 and Lemma 5.4 to see that

$$\begin{aligned} S_4(\tau) &= \sum_{\substack{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \\ \mathbf{0} < \mathbf{x}^{(j)} \leq \tau}} \Sigma_A(\{\mathbf{x}\}) \Sigma_B(\{\mathbf{x}\}; \mathbf{q}) \\ &= \mathbf{Q}^\gamma \sum_{\substack{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2r)} \in \mathbb{Z}^k \\ \mathbf{0} < \mathbf{x}^{(j)} \leq \tau}} \Xi(\mathbf{F}; \{\mathbf{x}\}) \Sigma_B(\{\mathbf{x}\}; \mathbf{q}) \\ &= \mathbf{Q}^\gamma \sum_{\mathcal{S} \subseteq \{1, \dots, k\}} \sum_{\{\mathbf{x}\} \in \mathcal{B}(\mathcal{S}; \tau)} \Xi(\mathbf{F}; \{\mathbf{x}\}) \Sigma_B(\{\mathbf{x}\}; \mathbf{q}) \\ (6.4) \quad &\ll \mathbf{Q}^\gamma \|\mathbf{q}\|^{1/2} \sum_{\mathcal{S} \subseteq \{1, \dots, k\}} N(\mathcal{S}; \tau) \mathbf{q}^{\sigma(\mathcal{S})/2}. \end{aligned}$$



We now bound  $N(\mathcal{S}; \tau)$  for each subset  $\mathcal{S}$ . If  $\mathcal{S} = \emptyset$ , the size of  $\mathcal{B}(\mathcal{S}; \tau)$  is no smaller than  $O(\tau_{\max}^{2rk})$ , so the key restriction comes from the fact that we are counting points that also lie in  $V_r(\mathbf{F}; \tau)$ . Thus for  $\mathcal{S} = \emptyset$  we use the estimate

$$N(\mathcal{S}; \tau) \leq \#V_r(\mathbf{F}; \tau) \leq J_r(\mathbf{F}; \tau_{\max}) \ll \tau_{\max}^{2rk-M+\varepsilon},$$

under the assumption that the bound (6.3) for  $J_r(\mathbf{F}; X)$  holds. For any non-empty  $\mathcal{S}$ , we use an upper bound based only on the size of  $\mathcal{B}(\mathcal{S}; \tau)$  from Lemma 5.3:

$$N(\mathcal{S}; \tau) \leq \#\mathcal{B}(\mathcal{S}; \tau) \ll \left( \prod_{i \in \mathcal{S}} \tau_i^r \right) \left( \prod_{i \notin \mathcal{S}} \tau_i^{2r} \right).$$

We then have

$$S_4(\tau) \ll \mathbf{Q}^\gamma \left\{ \tau_{\max}^{2rk-M+\varepsilon} \|\mathbf{q}\|^{\frac{1}{2}} + \|\mathbf{q}\|^{\frac{1}{2}} \sum_{\substack{\mathcal{S} \subseteq \{1, \dots, k\} \\ \mathcal{S} \neq \emptyset}} \left( \prod_{i \in \mathcal{S}} \tau_i^r \right) \left( \prod_{i \notin \mathcal{S}} \tau_i^{2r} \right) \mathbf{q}^{\sigma(\mathcal{S})/2} \right\}.$$

We henceforward assume that

$$(6.5) \quad r > M,$$

so that certainly  $2rk - M \geq 0$ . We now define  $K_{\max} = \max\{K_1, \dots, K_k\}$  and  $K_{\min} = \min\{K_1, \dots, K_k\}$  and use the fact that the above bound for  $S_4(\tau)$  is monotone in  $\tau \leq \mathbf{K}$ . Then  $\sup_{\tau \leq \mathbf{K}} S_4(\tau)$  is at most

$$\begin{aligned} & \ll \mathbf{Q}^\gamma \left\{ K_{\max}^{2rk-M+\varepsilon} \|\mathbf{q}\|^{\frac{1}{2}} + \|\mathbf{q}\|^{\frac{1}{2}} \sum_{\substack{\mathcal{S} \subseteq \{1, \dots, k\} \\ \mathcal{S} \neq \emptyset}} \left( \prod_{i \in \mathcal{S}} K_i^r \right) \left( \prod_{i \notin \mathcal{S}} K_i^{2r} \right) \mathbf{q}^{\sigma(\mathcal{S})/2} \right\} \\ & \ll \mathbf{Q}^\gamma \left\{ K_{\max}^{2rk-M+\varepsilon} \|\mathbf{q}\|^{\frac{1}{2}} + \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{\frac{1}{2}} \sum_{\substack{\mathcal{S} \subseteq \{1, \dots, k\} \\ \mathcal{S} \neq \emptyset}} \mathbf{K}^{-r\sigma(\mathcal{S})} \mathbf{q}^{\sigma(\mathcal{S})/2} \right\} \\ & = \mathbf{Q}^\gamma \left\{ K_{\max}^{2rk-M+\varepsilon} \|\mathbf{q}\|^{\frac{1}{2}} + \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{\frac{1}{2}} \sum_{\substack{\mathcal{S} \subseteq \{1, \dots, k\} \\ \mathcal{S} \neq \emptyset}} \left( \prod_{i \in \mathcal{S}} K_i^{-r} q_i^{1/2} \right) \right\}. \end{aligned}$$

Now we make the assumption that for every  $i = 1, \dots, k$  we have  $K_i^{-r} q_i^{1/2} \leq 1$ , or equivalently

$$(6.6) \quad K_i \geq q_i^{\frac{1}{2r}},$$

so that the largest contribution from the sum over subsets  $\mathcal{S}$  comes from sets of cardinality one. Then we have

$$\sup_{\tau \leq \mathbf{K}} S_4(\tau) \ll \mathbf{Q}^\gamma \left\{ K_{\max}^{2rk-M+\varepsilon} \|\mathbf{q}\|^{\frac{1}{2}} + \|\mathbf{K}\|^{2r} \|\mathbf{q}\|^{\frac{1}{2}} \sum_{i=1}^k K_i^{-r} q_i^{1/2} \right\}.$$

This implies

$$\sup_{\tau \leq \mathbf{K}} S_4(\tau) \ll \mathbf{Q}^\gamma \left\{ K_{\max}^{2rk+\varepsilon} K_{\min}^{-M} \|\mathbf{q}\|^{\frac{1}{2}} + K_{\max}^{2rk} \|\mathbf{q}\|^{\frac{1}{2}} \sum_{i=1}^k K_i^{-r} q_i^{1/2} \right\}.$$

The first term in braces dominates all other terms as long as for each  $i = 1, \dots, k$  we have

$$(6.7) \quad K_{\min}^{-M} \gg K_i^{-r} q_i^{1/2},$$

which is certainly implied by the condition

$$(6.8) \quad K_i^{r-M} \geq q_i^{1/2};$$

we note that this condition would also guarantee (6.6).

We have proved:

**Proposition 6.2.** *If  $r > M$  and (6.8) holds for each  $i = 1, \dots, k$ , then*

$$\sup_{\tau \leq \mathbf{K}} S_4(\tau) \ll \mathbf{Q}^\gamma K_{\max}^{2rk+\varepsilon} K_{\min}^{-M} \|\mathbf{q}\|^{\frac{1}{2}}.$$

This is the refinement of Proposition 4.5 that we sought.

**6.1. Proof of Theorem 2.1.** We will now input this bound for  $S_4(\tau)$  with the choice  $\mathbf{K} = 2\mathbf{H}/\mathbf{P}$  into Proposition 4.4, always with the specification that  $r > M$  and  $\mathbf{Q}$  is chosen as in (6.2). For each  $i = 1, \dots, k$  we choose  $P_i$  such that

$$\frac{1}{2} H_i q_i^{-\frac{1}{2(r-M)}} \leq P_i < H_i q_i^{-\frac{1}{2(r-M)}}.$$

With this choice, (6.8) is satisfied; we also have  $P_i \geq 1$  as long as  $H_i > q_i^{\frac{1}{2(r-M)}}$ . In order to satisfy the conditions  $H_i P_i < q_i$  of Lemma 3.2, we must also restrict to ranges with  $H_i < q_i^{\frac{1}{2} + \frac{1}{4(r-M)}}$ . With these choices, we apply Proposition 6.2 in Proposition 4.4 to obtain

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} \left\{ \mathbf{Q}^\gamma K_{\max}^{2rk+\varepsilon} K_{\min}^{-M} \|\mathbf{q}\|^{1/2} \right\}^{\frac{1}{2r}}.$$

We recall that

$$\mathbf{Q}^{\frac{\gamma}{2r}} \ll (\mathbf{H}/\mathbf{P})^{\frac{\gamma}{2r}} \ll (\mathbf{q}^\gamma)^{\frac{1}{4r(r-M)}}.$$

Thus we may conclude

$$\begin{aligned} T(\mathbf{F}; \mathbf{N}, \mathbf{H}) &\ll \|\mathbf{H}\|^{-\frac{1}{2r}} \|\mathbf{P}\|^{1-\frac{1}{2r}} (\mathbf{q}^\gamma)^{\frac{1}{4r(r-M)}} q_{\max}^{\frac{2rk}{4r(r-M)}} q_{\min}^{-\frac{M}{4r(r-M)}} \|\mathbf{q}\|^{\frac{1}{4r}+\varepsilon} \\ &\ll \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{-\frac{1}{2(r-M)}(1-\frac{1}{2r})} (\mathbf{q}^\gamma)^{\frac{1}{4r(r-M)}} q_{\max}^{\frac{2rk}{4r(r-M)}} q_{\min}^{-\frac{M}{4r(r-M)}} \|\mathbf{q}\|^{\frac{1}{4r}+\varepsilon} \\ &\ll \|\mathbf{H}\|^{1-\frac{1}{r}} \|\mathbf{q}\|^{\frac{-r-M+1}{4r(r-M)}+\varepsilon} (\mathbf{q}^\gamma)^{\frac{1}{4r(r-M)}} q_{\max}^{\frac{2rk}{4r(r-M)}} q_{\min}^{-\frac{M}{4r(r-M)}}, \end{aligned}$$

which proves Theorem 2.1. (Here we note that certainly  $r > R(d+1)$  implies  $r > M$ .)

In the case where  $q_i = q$  for all  $i$ , we have  $q_{\max} = q_{\min} = q$ ,  $\|\mathbf{q}\| = q^k$ , and  $\mathbf{q}^\gamma = q^M$  where  $M$  is the weight of the system  $\mathbf{F}$ , so that this simplifies to

$$T(\mathbf{F}; \mathbf{N}, \mathbf{H}) \ll \|\mathbf{H}\|^{1-\frac{1}{r}} q^{\frac{k(r+1-M)}{4r(r-M)}+\varepsilon}.$$

**6.2. Optimal choice of  $r$ .** Suppose that for each  $i$ ,  $H_i = q^{1/4+\kappa_i}$ . Set  $\sigma = \sum_{i=1}^k \kappa_i$ . Then Corollary 2.1.1 provides an upper bound of size  $\|H\|q^{-\delta+\varepsilon}$  where

$$\delta = \frac{4\sigma(r-M) - k}{4r(r-M)}.$$

As a function of  $r$ , this attains a maximum at the real value

$$r_0 = M + \frac{k \left( 1 + \sqrt{\frac{4M\sigma}{k} + 1} \right)}{4\sigma}.$$

Choosing  $r = r_0 + \theta$  where  $-1/2 \leq \theta < 1/2$ , we see that  $\delta$  is approximately of size

$$\delta \approx \frac{4\sigma^2}{k \left( 1 + \sqrt{\frac{4M\sigma}{k} + 1} \right)^2}.$$

For fixed  $k, d$  as  $\sigma = \sum \kappa_i \rightarrow 0$  this behaves like

$$\delta \approx \frac{\sigma^2}{k},$$

which we note is nicely dependent only on the dimension  $k$  and not on other parameters of the system  $\mathbf{F}$ .

## 7. TECHNICAL LEMMAS

**7.1. Proof of Lemma 3.2.** We now return to the proof of Lemma 3.2. It is clear from the definition of  $\mathcal{A}(\mathbf{m})$  that it vanishes unless each  $m_i$  satisfies  $|m_i| \leq 2q_i$ . Next we note that  $\mathcal{A}(\mathbf{m})$  is a non-negative integer, so trivially  $S_1 \leq S_2$ . Thus we turn to bounding  $S_2$ , for which we note that

$$\begin{aligned} \sum_{\mathbf{m}} \mathcal{A}(\mathbf{m})^2 &= \sum_{\mathbf{m}} \#\{\mathbf{p}, \mathbf{p}', \mathbf{a}, \mathbf{a}' : 0 \leq a_i < p_i, 0 \leq a'_i < p'_i \text{ and } p_i, p'_i \in \mathcal{P}_i : \\ &\quad m_i \leq \frac{N_i - a_i q_i}{p_i} < m_i + \frac{H_i}{P_i}, m_i \leq \frac{N_i - a'_i q_i}{p'_i} < m_i + \frac{H_i}{P_i}, i = 1, \dots, k\}. \end{aligned}$$

For a fixed  $\mathbf{m}$ , in order for a quadruple  $\mathbf{p}, \mathbf{p}', \mathbf{a}, \mathbf{a}'$  to belong to this set we must have both  $(N_i - a_i q_i)/p_i$  and  $(N_i - a'_i q_i)/p'_i$  belong to the interval  $[m_i, m_i + H_i/P_i)$  (for all  $i$ ), so that we require

$$\left| \frac{N_i - a_i q_i}{p_i} - \frac{N_i - a'_i q_i}{p'_i} \right| \leq \frac{H_i}{P_i}, \quad \text{for each } i = 1, \dots, k.$$

If these conditions are satisfied then there will be  $O(\|\mathbf{H}\| \|\mathbf{P}\|^{-1})$  corresponding values  $\mathbf{m}$  for which this can occur. We may thus deduce that

$$\begin{aligned} \sum_{\mathbf{m}} \mathcal{A}(\mathbf{m})^2 &\ll \|\mathbf{H}\| \|\mathbf{P}\|^{-1} \#\{\mathbf{p}, \mathbf{p}', \mathbf{a}, \mathbf{a}' : 0 \leq \left| \frac{N_i - a_i q_i}{p_i} - \frac{N_i - a'_i q_i}{p'_i} \right| \leq \frac{H_i}{P_i}\} \\ (7.1) \quad &\ll \|\mathbf{H}\| \|\mathbf{P}\|^{-1} \sum_{\mathbf{p}, \mathbf{p}' \in \mathcal{P}} \mathcal{M}(\mathbf{p}, \mathbf{p}'), \end{aligned}$$

where we set

$$\mathcal{M}(\mathbf{p}, \mathbf{p}') := \#\{\mathbf{0} \leq \mathbf{a} < \mathbf{p}, \mathbf{0} \leq \mathbf{a}' < \mathbf{p}' : \\ 0 \leq \left| \frac{N_i - a_i q_i}{p_i} - \frac{N_i - a'_i q_i}{p'_i} \right| \leq \frac{H_i}{P_i}, \text{ for each } i = 1, \dots, k\}.$$

We now define for any primes  $p_i, p'_i \in \mathcal{P}_i$  the quantity

$$M_i(p_i, p'_i) = \#\{0 \leq a_i < p_i, 0 \leq a'_i < p'_i : 0 \leq \left| \frac{N_i - a_i q_i}{p_i} - \frac{N_i - a'_i q_i}{p'_i} \right| \leq \frac{H_i}{P_i}\}.$$

We note that for each pair of tuples  $\mathbf{p}, \mathbf{p}'$ ,

$$\mathcal{M}(\mathbf{p}, \mathbf{p}') = \prod_{i=1}^k M_i(p_i, p'_i).$$

Thus

$$\begin{aligned} \sum_{\mathbf{p}, \mathbf{p}' \in \mathcal{P}} \mathcal{M}(\mathbf{p}, \mathbf{p}') &= \sum_{\mathbf{p}, \mathbf{p}' \in \mathcal{P}} \left( \prod_{i=1}^k M_i(p_i, p'_i) \right) \\ &= \prod_{i=1}^k \left( \sum_{p_i, p'_i \in \mathcal{P}_i} M_i(p_i, p'_i) \right) \\ (7.2) \quad &= \prod_{i=1}^k \left( \sum_{p_i = p'_i \in \mathcal{P}_i} M_i(p_i, p'_i) + \sum_{p_i \neq p'_i \in \mathcal{P}_i} M_i(p_i, p'_i) \right). \end{aligned}$$

We now recall from the proof of Lemma 2.2 in [13] that in the one-dimensional case it is already known that for each  $i = 1, \dots, k$ ,

$$\begin{aligned} \sum_{p_i \in \mathcal{P}_i} M_i(p_i, p_i) &\ll P_i^2 \\ \sum_{p_i \neq p'_i \in \mathcal{P}_i} M_i(p_i, p'_i) &\ll P_i^2, \end{aligned}$$

with the latter bound holding under the condition  $P_i H_i < q_i$ . Applying this in (7.2), we see that

$$\sum_{\mathbf{p}, \mathbf{p}' \in \mathcal{P}} \mathcal{M}(\mathbf{p}, \mathbf{p}') \ll \|\mathbf{P}\|^2,$$

so that in total (7.1) shows that

$$\sum_{\mathbf{m}} \mathcal{A}(\mathbf{m})^2 \ll \|\mathbf{H}\| \|\mathbf{P}\|,$$

as desired.

**7.2. Proof of Lemma 3.3.** Recall that in Lemma 3.3 we consider the sum

$$(7.3) \quad \sum_{\mathbf{n} \in I} a(\mathbf{n}),$$

for arbitrary complex numbers  $a(\mathbf{n})$  indexed by  $\mathbf{n} \in \mathbb{Z}^k$  lying in an arbitrary fixed product of sub-intervals  $I \subseteq (\mathbf{A}, \mathbf{A} + \mathbf{B}]$ . We will denote  $I = (\mathbf{C}, \mathbf{C} + \mathbf{D}]$ , with  $(C_i, C_i + D_i] \subseteq (A_i, A_i + B_i]$  for each  $i = 1, \dots, k$ . We note that if any  $D_i = 0$ , then the sum (7.3) is vacuous; thus we may assume all  $D_i > 0$ . Next note that if  $B_i < 1$  then there is at most one value  $n_i$  considered in the  $i$ -th coordinate of the sum  $\sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n})$ , and we could regard the sum as living in a lower dimensional setting and proceed with the proof in a lower dimension. Thus we may assume  $B_i \geq 1$  for all  $i = 1, \dots, k$ .

We will prove Lemma 3.3 with a simple adaptation of Bombieri and Iwaniec's original argument [1]. For each  $i$ , let  $\psi_i(x)$  denote a  $C^\infty$  compactly supported non-negative function that vanishes for  $x \leq \lfloor C_i \rfloor$  and  $x \geq \lfloor C_i + D_i \rfloor + 1$  and is identically 1 for  $\lfloor C_i \rfloor + 1 \leq x \leq \lfloor C_i + D_i \rfloor$ ; clearly we may also choose this so that  $|\psi_i| \leq 1$  and  $\psi_i$  has uniformly bounded derivatives  $|\psi_i^{(N)}| \leq 1$  for all  $N \geq 1$ . Let  $\Psi(\mathbf{x}) = \psi_1(x_1) \cdots \psi_k(x_k)$ , so that

$$\sum_{\mathbf{n} \in (\mathbf{C}, \mathbf{C} + \mathbf{D}]} a(\mathbf{n}) = \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n}) \Psi(\mathbf{n}).$$

After expressing  $\Psi(\mathbf{x})$  in terms of its inverse Fourier transform (see (7.6)), we have

$$\begin{aligned} \sum_{\mathbf{n} \in (\mathbf{C}, \mathbf{C} + \mathbf{D}]} a(\mathbf{n}) &= \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n}) \int_{\mathbb{R}^k} \hat{\Psi}(\boldsymbol{\theta}) e(\mathbf{n} \cdot \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^k} \hat{\Psi}(\boldsymbol{\theta}) \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n}) e(\mathbf{n} \cdot \boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Thus

$$(7.4) \quad \left| \sum_{\mathbf{n} \in (\mathbf{C}, \mathbf{C} + \mathbf{D}]} a(\mathbf{n}) \right| \leq \sup_{\boldsymbol{\theta} \in \mathbb{R}^k} \left| \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B}]} a(\mathbf{n}) e(\mathbf{n} \cdot \boldsymbol{\theta}) \right| \|\hat{\Psi}\|_{L^1(\mathbb{R}^k)},$$

where

$$\|\hat{\Psi}\|_{L^1(\mathbb{R}^k)} = \int_{\mathbb{R}^k} |\hat{\Psi}(\boldsymbol{\theta})| d\boldsymbol{\theta}.$$

We now note that

$$\hat{\Psi}(\boldsymbol{\theta}) = \int_{\mathbb{R}^k} \Psi(\mathbf{x}) e(-\mathbf{x} \cdot \boldsymbol{\theta}) d\mathbf{x} = \prod_{i=1}^k \left( \int_{A_i-1}^{A_i+B_i+1} \psi_i(x_i) e(-x_i \theta_i) dx_i \right) = \prod_{i=1}^k J_i(\theta_i),$$

say, where we define

$$J_i(\theta) = \int_{A_i-1}^{A_i+B_i+1} \psi_i(x) e(-x\theta) dx.$$

Each of these satisfies

$$(7.5) \quad |J_i(\theta)| \ll \min\{B_i + 2, |\theta|^{-N}\}, \quad \text{for any } N \geq 1.$$

The first option is the trivial bound; the second option follows from integration by parts. Precisely, for a fixed  $\theta \neq 0$ , by writing

$$e^{-2\pi i \theta x} = \frac{1}{(-2\pi i \theta)} \frac{d}{dx} e^{-2\pi i \theta x},$$

we may integrate by parts repeatedly to see that

$$\int_{A_i-1}^{A_i+B_i+1} \psi_i(x) e(-x\theta) dx = \frac{1}{(2\pi i\theta)^N} \int_{A_i-1}^{A_i+B_i+1} \psi_i^{(N)}(x) e(-x\theta) dx$$

for any  $N \geq 1$ ; the boundary terms vanish due to the compact support of  $\psi_i$ . Now we note that  $\psi_i^{(N)}(x)$  is uniformly bounded by assumption, and moreover it vanishes unless  $x$  belongs to either of two intervals of length 1. This gives the desired result (7.5).

We temporarily set  $L_i = B_i + 2$  and apply (7.5) to observe that for each  $i$ :

$$\begin{aligned} \int_{-\infty}^{\infty} |J_i(\theta)| d\theta &\ll \int_{|\theta| \leq L_i^{-1}} L_i d\theta + \int_{L_i^{-1} \leq |\theta| \leq 2L_i} |\theta|^{-1} d\theta + \int_{|\theta| \geq 2L_i} |\theta|^{-2} d\theta \\ &\ll 1 + \log(L_i) + L_i^{-1} \ll \log(B_i + 2). \end{aligned}$$

Finally, we see that

$$\int_{\mathbb{R}^k} |\hat{\Psi}(\boldsymbol{\theta})| d\boldsymbol{\theta} = \prod_{i=1}^k \left( \int_{-\infty}^{\infty} |J_i(\theta_i)| d\theta_i \right) \ll \prod_{i=1}^k \log(B_i + 2),$$

confirming that

$$(7.6) \quad \hat{\Psi} \in L^1(\mathbb{R}^k).$$

Thus the use of the Fourier inversion formula is justified, and we also see in (7.4) that

$$\left| \sum_{\mathbf{n} \in (\mathbf{C}, \mathbf{C} + \mathbf{D})} a(\mathbf{n}) \right| \ll \left( \prod_{i=1}^k \log(B_i + 2) \right) \sup_{\boldsymbol{\theta} \in \mathbb{R}^k} \left| \sum_{\mathbf{n} \in (\mathbf{A}, \mathbf{A} + \mathbf{B})} a(\mathbf{n}) e(\mathbf{n} \cdot \boldsymbol{\theta}) \right|,$$

as claimed.

**7.3. Proof of Lemma 4.2.** We will proceed by iterated partial summation applied to

$$(7.7) \quad \sum_{\mathbf{0} < \mathbf{n} \leq \mathbf{N}} a(\mathbf{n}) b(\mathbf{n}).$$

We first apply partial summation with respect to  $n_1$  in (7.7). We set  $J = \{1\}$  and  $I = \{2, \dots, k\}$  so that

$$\begin{aligned} \sum_{\mathbf{n} \leq \mathbf{N}} a(\mathbf{n}) b(\mathbf{n}) &= \sum_{\mathbf{n}_{(I)} \leq \mathbf{N}_{(I)}} \left( \sum_{0 < n_1 \leq N_1} a(n_1, \mathbf{n}_{(I)}) b(n_1, \mathbf{n}_{(I)}) \right) \\ &= \sum_{\mathbf{n}_{(I)} \leq \mathbf{N}_{(I)}} \left\{ b(N_1, \mathbf{n}_{(I)}) \left( \sum_{0 < n_1 \leq N_1} a(n_1, \mathbf{n}_{(I)}) \right) \right. \\ &\quad \left. - \int_0^{N_1} \left( \sum_{0 < n_1 \leq t_1} a(n_1, \mathbf{n}_{(I)}) \right) \frac{d}{dt_1} b(t_1, \mathbf{n}_{(I)}) dt_1 \right\}. \end{aligned}$$

We may then apply partial summation with respect to  $n_2$ , and so on, iteratively for each  $n_i$  with  $i \leq k$ . One obtains a representation of (7.7) as a sum of  $2^k$  terms, each corresponding

to a subset  $J \subseteq \{1, \dots, k\}$  (and its corresponding complement  $I$ ). For each partition  $J \cup I$  of  $\{1, \dots, k\}$  with  $|J| = v$ , the resulting term is of the shape

$$(-1)^v \int \cdots \int_{\substack{(0, N_j] \\ j \in J}} A_{(I), (J)}(\mathbf{N}_{(I)}, \mathbf{t}_{(J)}) \frac{\partial^v}{\partial \mathbf{t}_{(J)}} b(\mathbf{N}_{(I)}, \mathbf{t}_{(J)}) d\mathbf{t}_{(J)}.$$

Here if  $J = \{j_1, \dots, j_v\}$  we let  $\frac{\partial^v}{\partial \mathbf{t}_{(J)}} = \frac{\partial^v}{\partial t_{j_1} \cdots \partial t_{j_v}}$  and  $d\mathbf{t}_{(J)} = dt_{j_1} \cdots dt_{j_v}$ . As a result of the assumed bounds (4.10) on the partial derivatives of  $b(\mathbf{x})$ , we may conclude that

$$\left| \sum_{\mathbf{n} \leq \mathbf{N}} a(\mathbf{n}) b(\mathbf{n}) \right| \leq \sum_{J \subseteq \{1, \dots, k\}} \left( \prod_{j \in J} B_j \right) \int \cdots \int_{\substack{(0, N_j] \\ j \in J}} |A_{(cJ), (J)}(\mathbf{N}_{(cJ)}, \mathbf{t}_{(J)})| d\mathbf{t}_{(J)},$$

which is the statement of the lemma.

**Acknowledgments.** The author is supported in part by NSF DMS-1402121, and also thanks the Hausdorff Center for Mathematics for a very pleasant working environment during a portion of this research. The author thanks D. R. Heath-Brown and T. D. Wooley for informative discussions on the work of this paper, D. Schindler for a close reading of an earlier version of the manuscript, and the anonymous referees for helpful comments on the exposition.

## REFERENCES

1. E. Bombieri and H. Iwaniec, *On the order of  $\zeta(1/2 + it)$* , Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **13** (1986), 449–472.
2. J. Bourgain and M.-C. Chang, *On a multilinear character sum of Burgess*, C. R. Acad. Sci. Paris, Ser. I **348** (2010), 115–120.
3. D. A. Burgess, *The distribution of quadratic residues and non-residues*, Mathematika **4** (1957), 106–112.
4. ———, *On character sums and  $L$ -series*, J. Reine Angew. Math. **3** (1962), 193–206.
5. ———, *On character sums and  $L$ -series II*, Proc. London Math. Soc. **3** (1963), 524–536.
6. ———, *A note on character sums of binary quadratic forms*, J. London Math. Soc. **43** (1968), 271–274.
7. ———, *The character sum estimate with  $r = 3$* , J. London Math. Soc. (2) **33** (1986), 219–226.
8. M.-C. Chang, *Burgess inequality in  $\mathbb{F}_{p^2}$* , Geom. Funct. Anal. **19** (2009), 1001–1016.
9. M.-C. Chang and I. E. Shparlinski, *On the density of integer points on generalised Markoff-Hurwitz and Dwork hypersurfaces*, arXiv:1408.4514 (2014).
10. P. X. Gallagher and H. L. Montgomery, *A note on Burgess’s estimate*, Mathematical Notes **88** (2010), 321–329.
11. D. R. Heath-Brown, *The least square-free number in an arithmetic progression*, J. Reine Angew. Math. **332** (1982), 204–220.
12. D. R. Heath-Brown, *Burgess’s bounds for character sums*, Proceedings in Mathematics and Statistics, Springer, New York **43** (2012), 199–213.
13. D. R. Heath-Brown and L. B. Pierce, *Burgess bounds for short mixed character sums*, J. London Math. Soc. (2) **91** (2015), 693–708.
14. S. Parsell, S. M. Prendiville, and T. D. Wooley, *Near-optimal mean value estimates for multidimensional Weyl sums*, Geom. Funct. Anal. **23** (2013), 1962–2024.
15. L. B. Pierce, *The 3-part of class numbers of quadratic fields*, J. London Math. Soc. **71** (2005), 579–598.
16. I. E. Shparlinski, *On the distribution of points on the generalized Markoff-Hurwitz and Dwork hypersurfaces*, International Journal of Number Theory **10** (2014), 151–160.

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY, 120 SCIENCE DRIVE, DURHAM NC 27708  
*E-mail address:* pierce@math.duke.edu